Experimental Techniques

Rick Van Kooten

Indiana University

Fifth CERN-Fermilab Hadron Collider Physics Summer School

Fermilab, Batavia, IL

24–26 Aug. 2010

Introduction

"Experimental Techniques"

Often taken as covering tracking, calorimetery, particle ID, triggering/DAQ, etc.

→ already covered

"Experimental Techniques"

Can often mean statistical methods as applied to data analysis & interpretation

→ already covered, Barlow

(and also realize how tough this can be, and what fraction of your time you may be dealing with it!)

"Experimental Techniques": doing a data analysis plus "filling in the gaps" of important items not yet covered, assemblage of examples and "how to's"

Outline

"Experimental Techniques" in the context of three quite very different types of analyses, seguing into topics important for that kind of analysis

"Absolute", e.g., measuring a cross section $\sigma(p\bar{p} \to Z^0 X) \cdot \mathcal{B}(Z^0 \to \mu^+ \mu^-)$

- Instantaneous & integrated luminosity (see Prebys talk for getting there)
- Triggers (efficiency & combining) (for rest see Vachon's talk)
- Efficiency / acceptance
- Monte Carlo simulations
- Unfolding

Measuring particle properties: e.g., B_s^0 lifetime

Top quark mass

W mass

- High p_T b-jet tagging
- Different ways to extract from observables
- Blind analyses
- Systematic Uncertainties

Outline

Searches for new particles/phenomena

Event selection

Multivariate Techniques

Backgrounds

Limits

Subtopics easily move back and forth among these different classes

Guaranteed that there are people here more expert than I am in many of these areas that I am! (That is what a Ph.D. or senior grad student is by definition!)

Glean what you can in areas that you have not worked in yet

Acknowledgements

Past lectures, e.g., Heinemann, Hoecker, etc. from who I have borrowed some material liberally

Preamble

The data you are analysing come from real detectors.
Since a real detector is not perfect (or because of basic physics reasons), the measurements have limitations:

- Do not measure all events/particles

  Finite acceptance (geometrical, kinematical)

- Cannot measure the true variable with infinite accuracy

  Finite resolution

- Cannot uniquely identify all events/particles

  Have to know detection/identification efficiency, purity, backgrounds

- Cannot uniquely identify underlying processes of event, or want to extract only specific subset of events

  Event selection (with again efficiency, backgrounds)

Measuring a Cross Section

...or any other "absolute" measurement...

Number of observed candidates
(fitted or counted)

Number of background candidates
(measured from data
or calculated from theory)
(minimize)

Cross section in cm^2
(or □b, nb, pb)

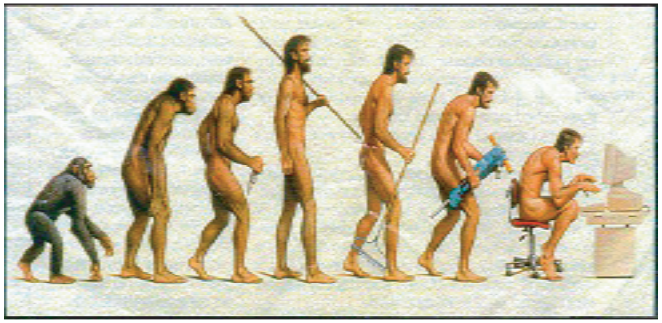
$$\sigma = \frac{N_{\mathrm{obs}} - N_{\mathrm{backg}}}{\epsilon \cdot \int \mathcal{L} dt}$$

Efficiency/acceptance
(maximize)

Integrated Luminosity in cm^{-1}
(or □b^{-1}, nb^{-1}, pb^{-1})
(maximize,
unless systematically limited)

$$\frac{\delta\sigma}{\sigma} = \sqrt{\frac{\delta N_{\mathrm{obs}}^2 + \delta N_{\mathrm{backg}}^2}{(N_{\mathrm{obs}} - N_{\mathrm{backg}})^2} + \left(\frac{\delta\mathcal{L}}{\mathcal{L}}\right)^2 + \left(\frac{\delta\epsilon}{\epsilon}\right)^2}$$

Typical Access to Data



Raw Data

Centrally managed reconstruction –
batch-like on farms/Grid, only once ideally

Reconstructed Data

Skimming – copying subsets of data,
usually different for different physics working groups

Skim dataset

Compress/subset of information, possibly after re-reconstruction

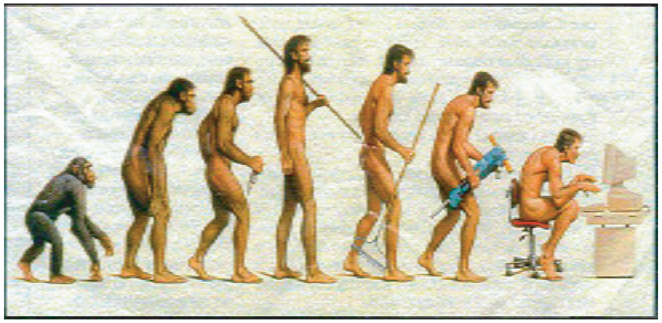
Analysis dataset(s)

What one regularly works on, "pre-selected" with loose selection criteria

- Small enough to run over and over with rapid turn-around
- Large enough to enable background estimation
- Try to retain clear parentage (so can determine luminosity, trigger effic.)
- Use standard, approved definitions of objects unless a good reason not to

Other

Typical Access to Data



Raw Data

How much is *my* analysis using??

Centrally managed reconstruction –
batch-like on farms/Grid, only once ideally

Reconstructed Data

Skimming – copying subsets of data,
usually different for different physics working groups

Skim dataset

Compress/subset of information, possibly after re-reconstruction

Analysis dataset(s)

What one regularly works on, "pre-selected" with loose selection criteria

Small enough to run over and over with rapid turn-around

Large enough to enable background estimation

Try to retain clear parentage (so can determine luminosity, trigger effic.)

Use standard, approved definitions of objects unless a good reason not to

Other

Measuring a Cross Section

Cross section in cm^2
(or □b, nb, pb)

$$\sigma = \frac{N_{\rm obs} - N_{\rm backg}}{\epsilon \cdot \int \mathcal{L} dt}$$

Integrated Luminosity in cm^{-2}
(or □b^{-1}, nb^{-1}, pb^{-1})
(maximize,
unless systematically limited)

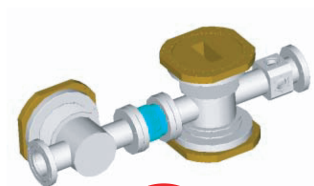
Measuring Luminosity

Lots of ways to measure it:

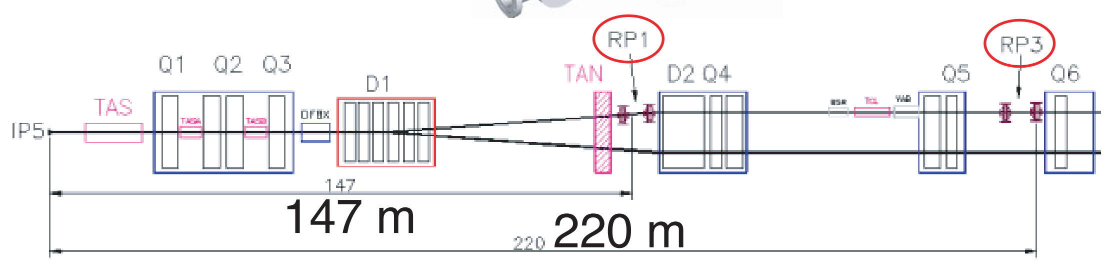
- Machine beam optics,
  estimate to ~20 – 30%

- Relate number of interactions to total cross section,
  absolute precision ~4-6%, relative precision much better

- Elastic *pp* cross section,
  tiny angles, "Roman Pots" ~few 100 m either side of
  interaction point, LHC expects absolute precision ~3%



Retract when injecting beam,
once colliding, insert to within
1 mm (!) of beam

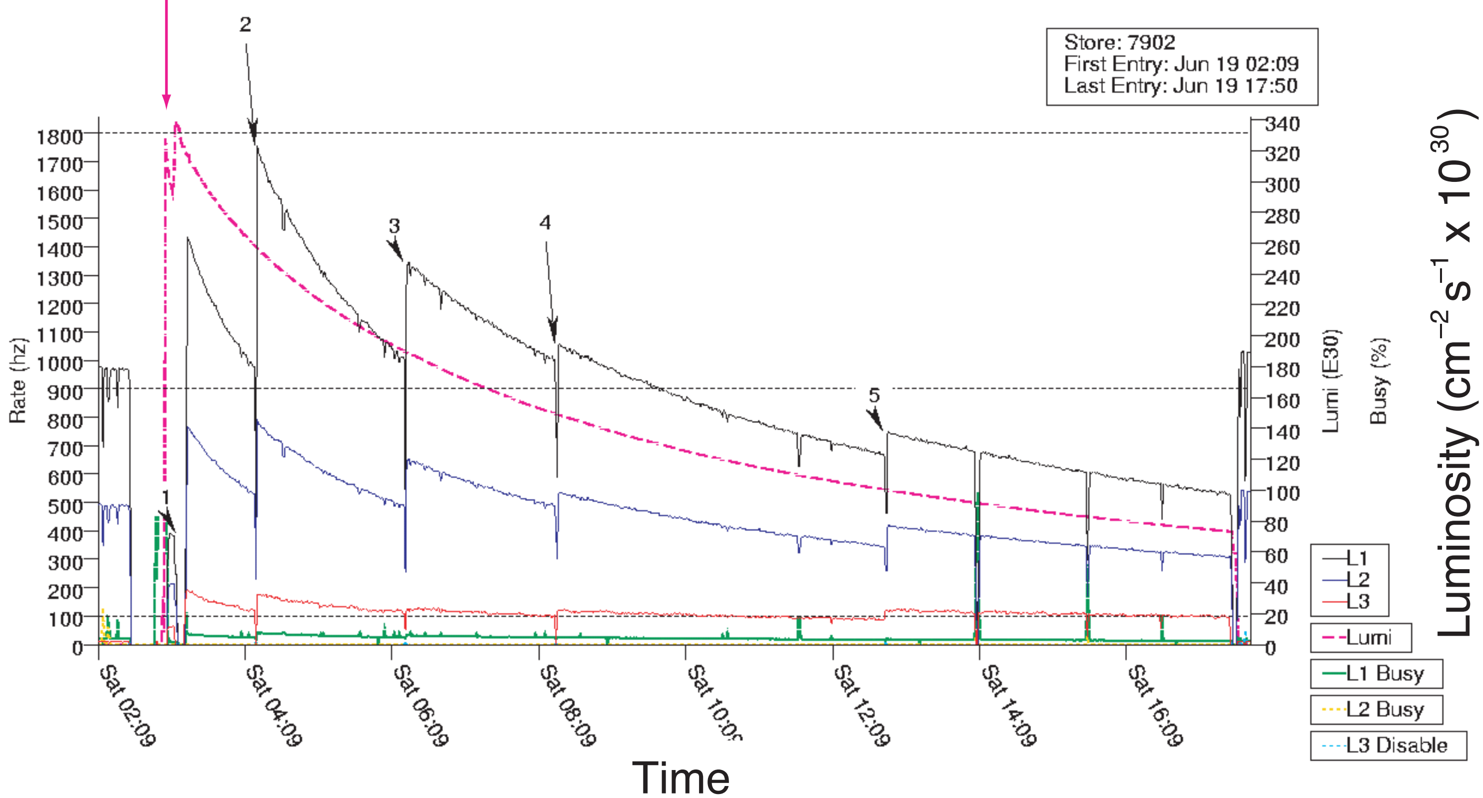


- Electroweak "candles", well-known processes,
  W and/or *Z* production, possible precision ~2–3%?

Measuring Luminosity

Need absolute number plus relative with time, fast measurement:

- $\mathcal{L} = \mathcal{L}_0 \exp(-t/\tau)$ instantaneous luminosity falls decays away with time

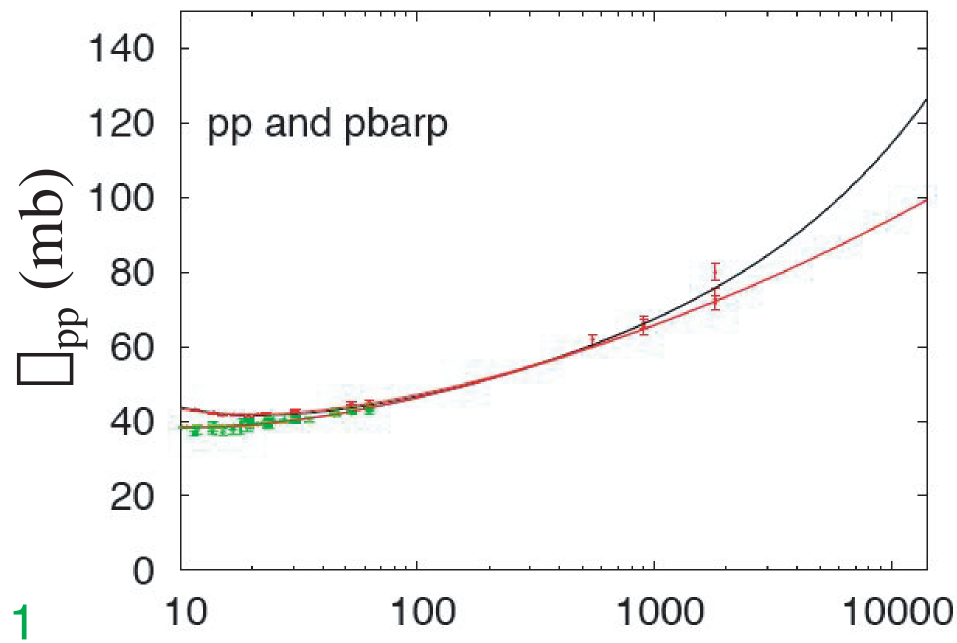


Measuring Luminosity

Rate of *pp* interactions: $R_{pp} = \sigma_{\text{inelastic}} \cdot \epsilon \cdot \mathcal{L}_{\text{inst}}$

Instantaneous

- Measure fraction of beam crossings *without* interactions
  - related to R_{pp}
- Relative normalization possible
  - if decent probability for no interactions, i.e.,
  - $\mathcal{L} < 10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$
- Absolute normalization
  - Normalize to measured inelastic *pp* cross section
  - Measured by CDF and E710/E811

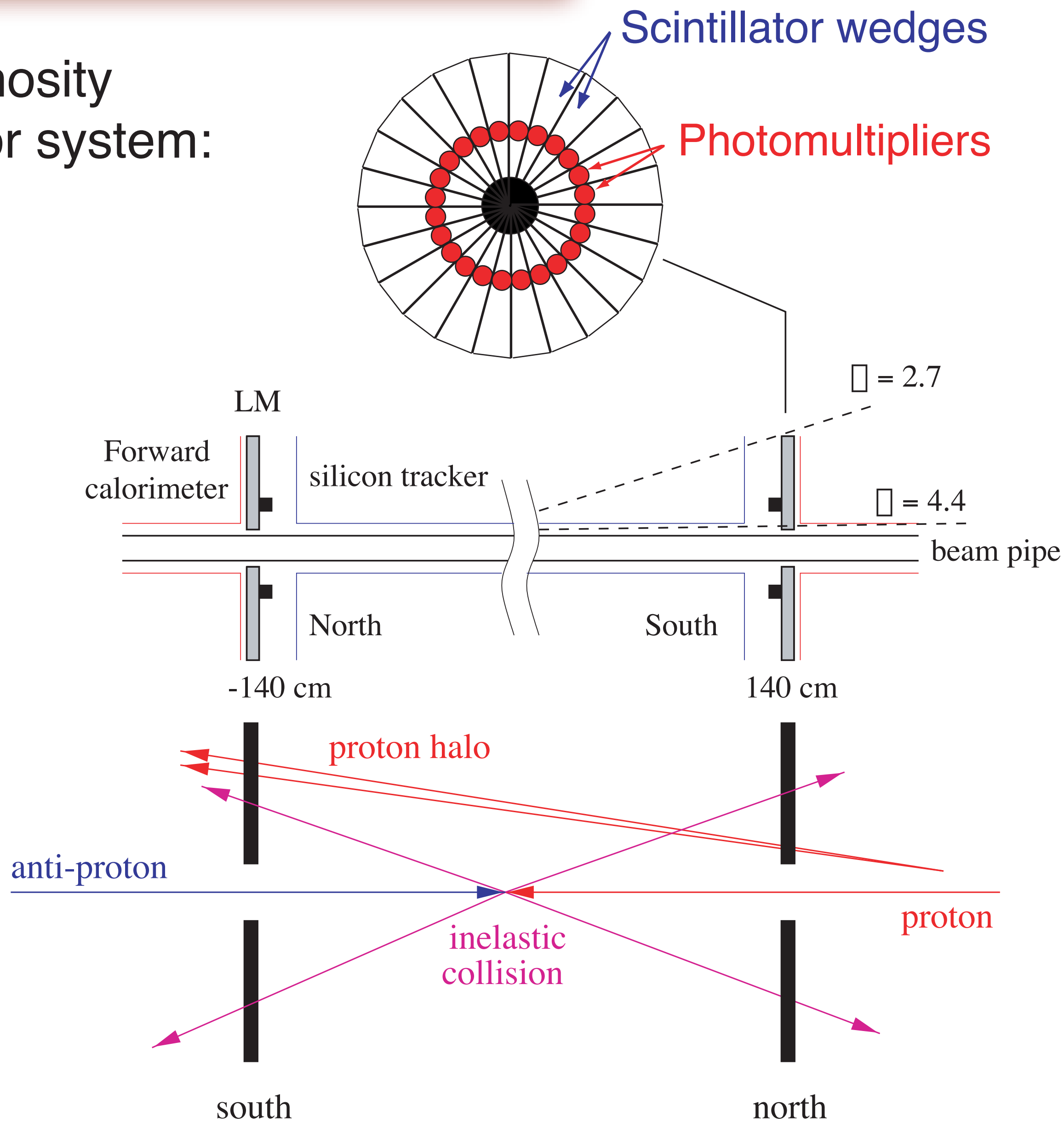


$$\sigma_{\text{inelastic}} = 60.7 \pm 2.4\,\text{mb} \quad \text{@ 1.96 TeV}$$

$$125 \pm 25\,\text{mb} \quad \text{@ 14 TeV}$$

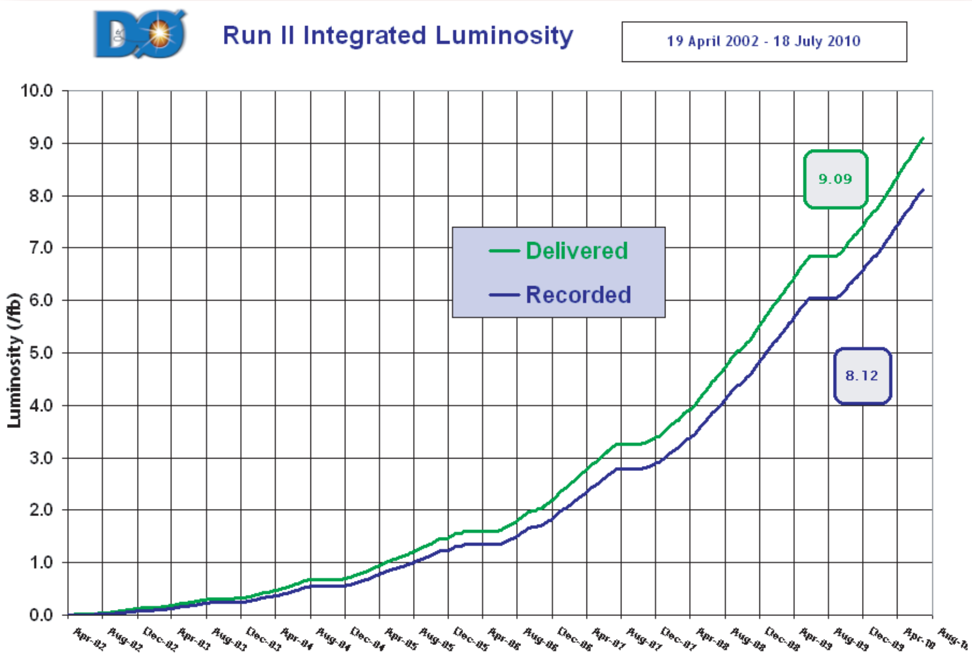
Measuring Luminosity

e.g., DØ luminosity monitor system:

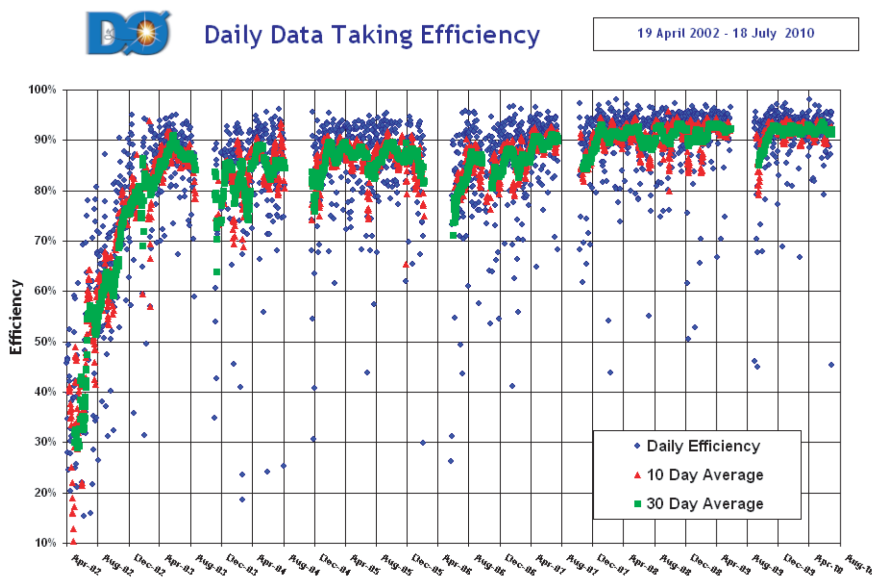


Measuring Luminosity

Luminosity [fb]



100%
90%
80%
Efficiency



...but delivered
luminosity π collected luminosity!

- Detector/shifters not 100% efficient
- Your trigger(s) may have been off or prescaled (described later) at some given time
- Some parts of the detector may not always be on or operational

Apply "data-quality" cuts at top level of analysis for sub-detectors you care about: e.g.,

Muon system ok?
Tracking systems ok?
Calorimeters ok?
May not need all of them!

Measuring Luminosity

This can be/is a bookkeeping nightmare!

Trust your colleagues/experts! Follow their recommended procedures and use their tools $\longrightarrow$ they have worried more about it, and it is often even worse of a bookkeeping nightmare than you imagine!

Sum up to get integrated luminosity:

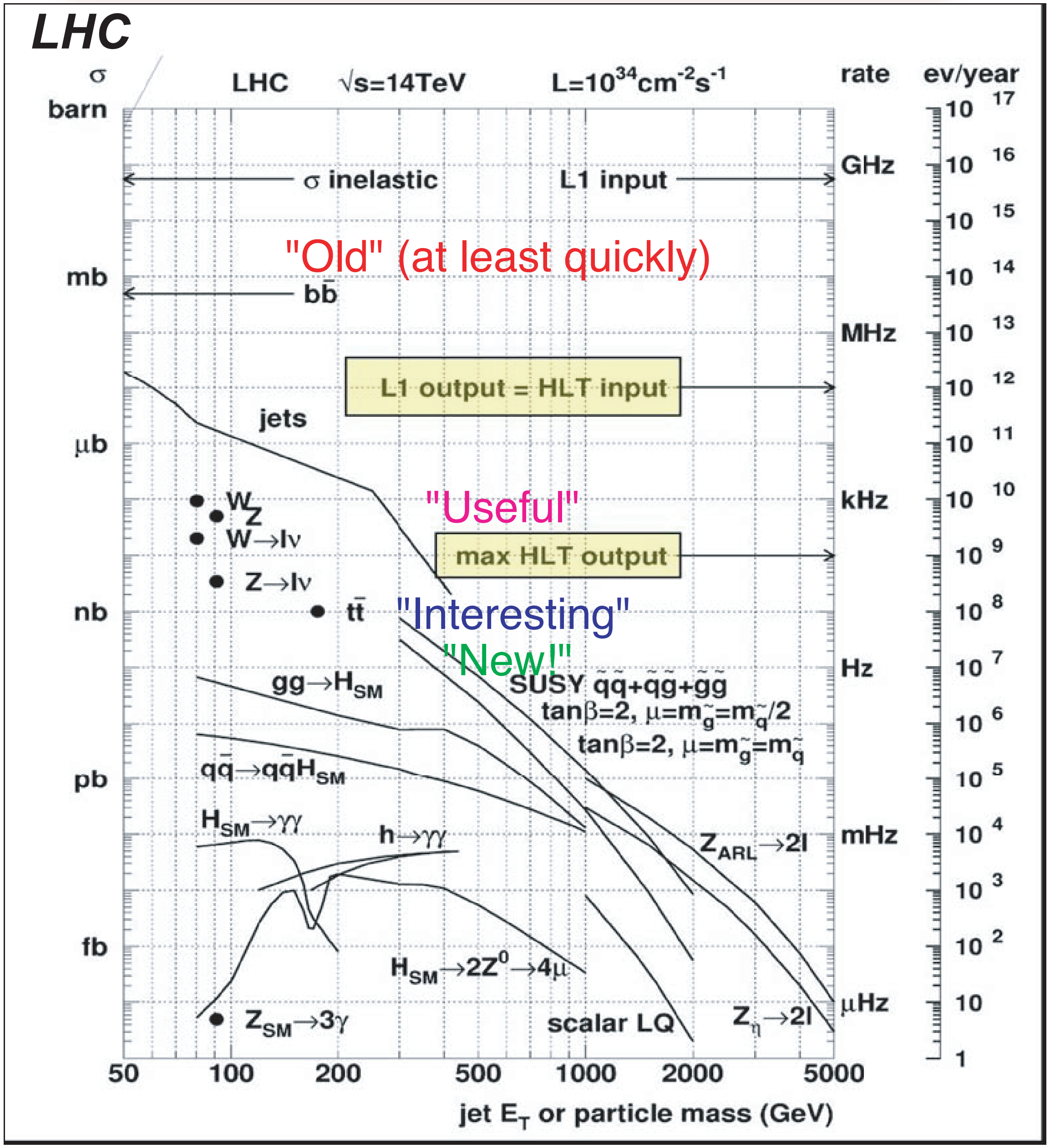
$$\int \mathcal{L} dt = \sum \mathcal{L}_{\mathrm{instant},i}(t, \ldots) \cdot \Delta t$$

Instantaneous Luminosity

Small chunk of time where your collision event(s) falls

Comes with an overall, absolute scale uncertainty ("luminosity constant") usually determined by others and usually broken out as a separate uncertainty: $\pm(\int \mathcal{L} dt)$

Triggering



Why? (Reminder)

See *Trigger/DAQ* by Vachon

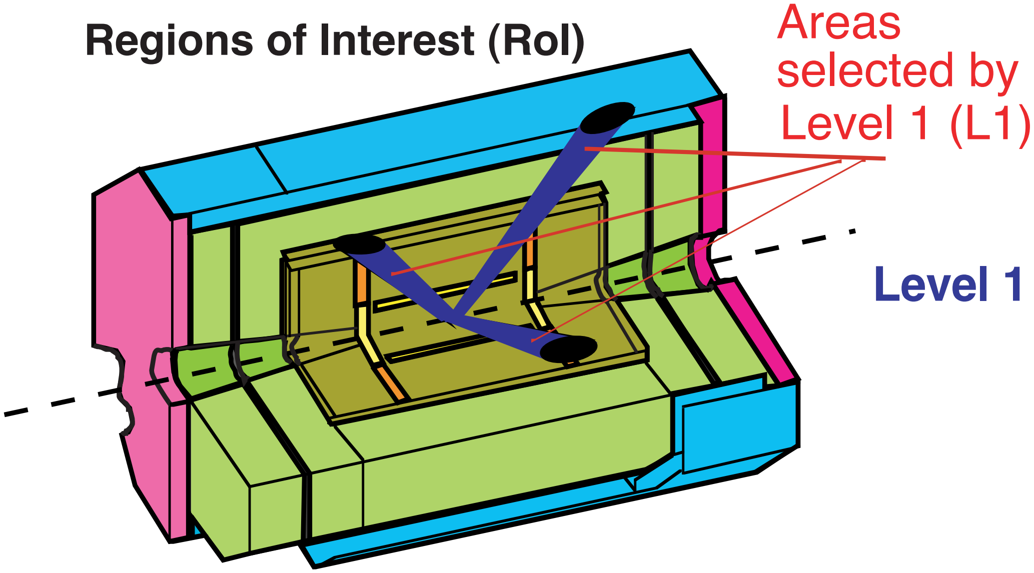
- Cannot (and do not want to) store all events; "interesting/useful/new physics" buried under "old physics"
- Look at (almost) all bunch crossings, select most interesting ones, collect all detector and store it (@ ~100 – 200 Hz, similar at Tevatron) for later offline analysis
- "Interesting/new physics" occurs mostly at rates of 10, 1, or < 0.1 Hz

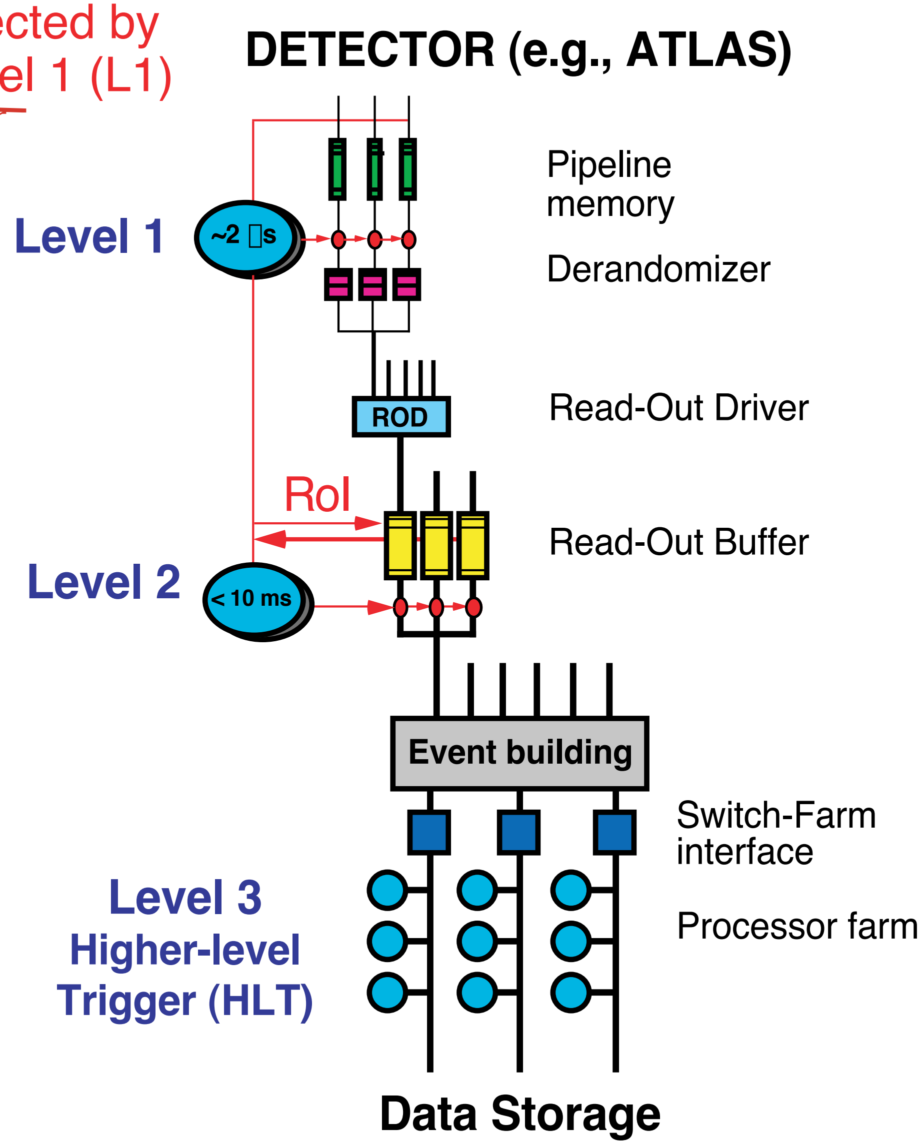
*Want to keep **all** these, reject **most** of the others*

Triggering

How? (Reminder)

See *Trigger/DAQ* by Vachon





Triggering

LEP: e^+e^- collider

- CM energy ~ 200 GeV
- Peak $L = 10^{32}$ cm^{-2}s^{-1}
- BC period: 22 μs
  (bunch-crossing,
  an ~eternity! triggering not tough,
  although *B* factories 4 – 8 ns!)

Tevatron: $p\bar{p}$ collider

- CM energy ~ 2 TeV
- $L = 3.5 \times 10^{32}$ cm^{-2}s^{-1}
- BC period: 396 ns

LHC: *pp* collider

- CM energy 14 TeV
- $L = 10^{34}$ cm^{-2}s^{-1}
- BC period: 25 ns

In e^+e^- colliders, interaction rate is very small compared to bunch-crossing rate (due to low cross-section)

- LEP/HERA: Usually selected events contain just a single interaction

LHC at design luminosity, each bunch-crossing will on average contain about 25 interactions! (and not too far from that at start of store of Tevatron)

- Your funky new physics event is recorded along with ~25 other proton-proton interactions
- These other interactions = "minimum-bias" interactions, i.e., the ones that would have been selected by a trigger that selects interactions in an (almost) unbiased way

HC Analysis → tough to trigger on, have to deal with the mess of all these other events ("pile up")

Triggering & Analysis

One of very first steps in analysis:
are the events that you are interested in being triggered??

Usually (trigger experts composing "trigger menus" are smart!)
particularly if event contains, e.g.:

- high-p_T leptons (or isolated leptons)
- multiple leptons
- large missing E_T
- multiple jets + something else...

Maybe not (or not efficiently), e.g.:

- low-momentum objects, (although lower efficiency may be okay, e.g., low-p_T B physics with cross section)
- to increase efficiency, may need to combine multiple triggers

If not → design one!

- (and fight for trigger bandwidth!)

Triggering & Analysis

Two key things you need for analysis:

- Prescales
- Trigger efficiency

$$\sigma = \frac{N_{\text{obs}} - N_{\text{backg}}}{\epsilon \cdot \int \mathcal{L} dt}$$

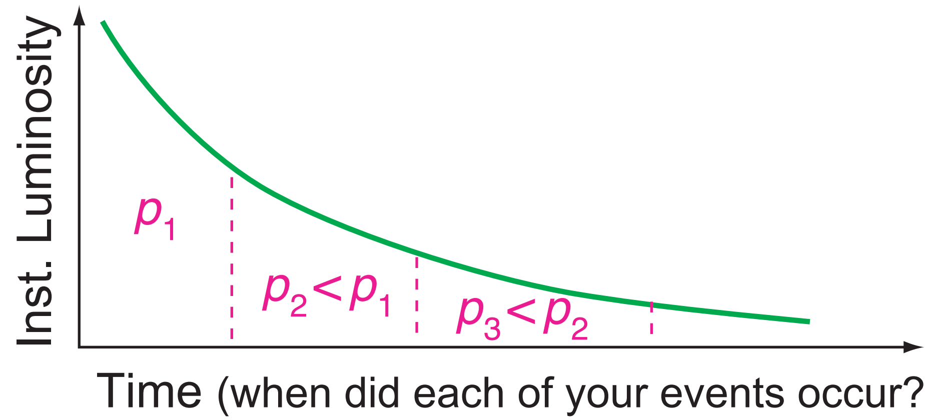
$$\epsilon_{\text{(total)}} = \epsilon_{\textbf{trig}} \cdot \epsilon_{\text{rec}} \cdot \epsilon_{\text{ID}} \cdot \epsilon_{\text{kin}}$$

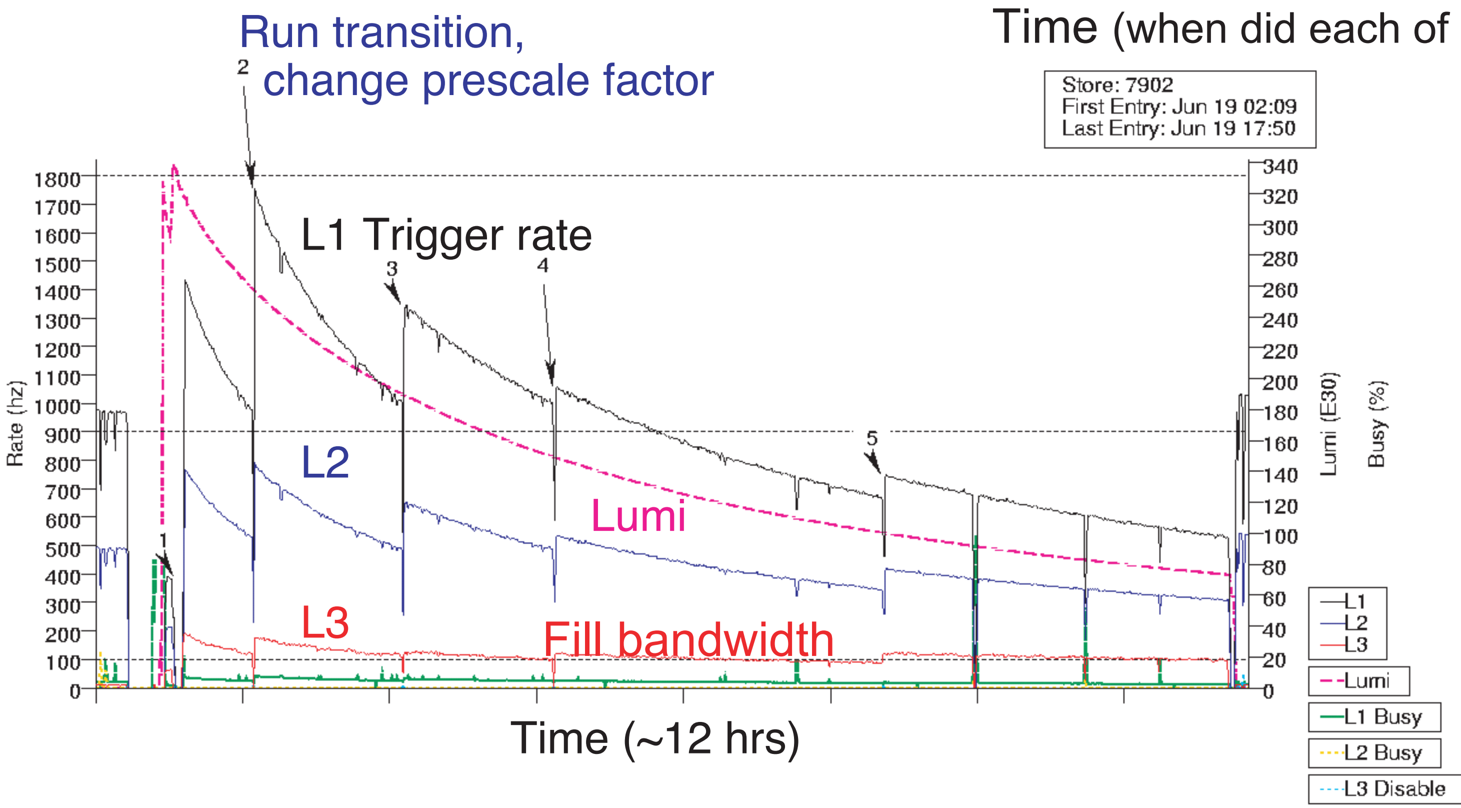
Wonders of Prescales

Not possible to keep all triggers at high luminosity and/or want to monitor types of events

→ take every *p*th event, where *p* = prescale factor

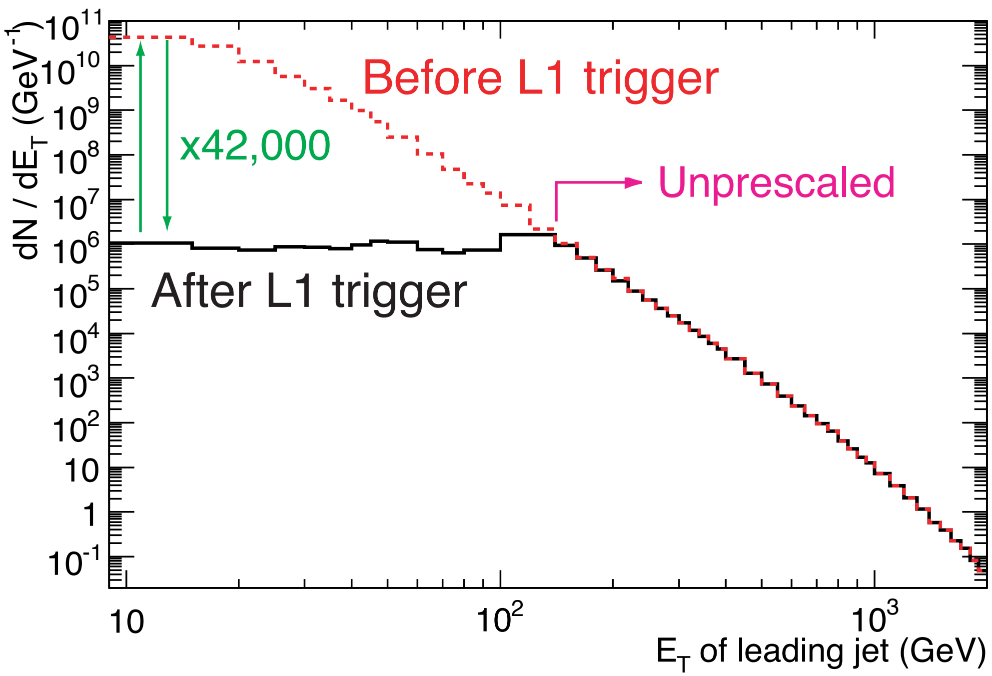
- Prescale as a function of time:





Wonders of Prescales

- Prescale as a function trigger requirement:



CERN-OPEN-2008-020; Hoecker

Trigger	Overall Prescale	Rate (Hz)
4j23	1	6.9 (± 0.8)
4j18	100	0.14 (± 0.01)
4j10	300	0.045 (± 0.004)
3j18	100	0.92 (± 0.03)
3j10	1500	0.061 (± 0.002)
Total Multi-Jets		7.9 (± 0.2)
j120	1	8.7 (± 0.9)
j70	15	4.2 (± 0.2)
j42	100	3.73 (± 0.06)
j35	500	1.37 (± 0.02)
j23	2000	1.37 (± 0.008)
j18	6000	1.02 (± 0.004)
j10	42000	3.9 (± 0.003)
j5	300000	0.9470 (± 0.0004)
Total Single-Jets		24.40 (± 0.01)

Expected rate out of ATLAS High-Level trigger
at 10^{31} $cm^{-2}s^{-1}$ peak luminosity
(low lumi at start up)

e.g., j10 = (E_T of leading jet) > 10 GeV

Wonders of Prescales

Want mostly *unprescaled triggers* for primary physics goals; examples:

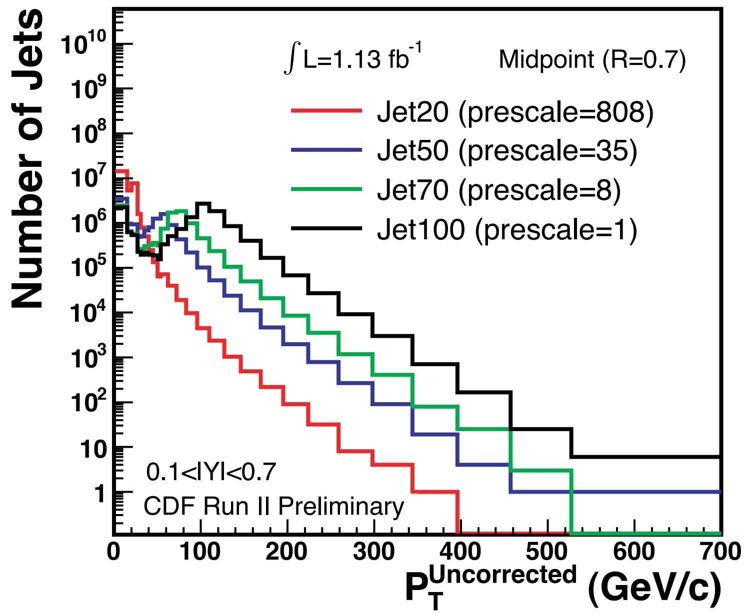
	ATLAS(*) (L=2x10^{33} cm^{-2}s^{-1})		CDF (L=3x10^{32} cm^{-2}s^{-1})
MET	e.g., SUSY, ...	> 70 GeV	> 40 GeV
Jet	jet s, monojets	> 370 GeV	> 100 GeV
Photon (iso)	GMSB SUSY Jet energy scale	> 55 GeV	> 25 GeV
Muon	*W*, *Z*, top, *WH* SUSY, *W'*, *Z'*	iso + p_T> 20 GeV	> 20 GeV
Electron	"	Iso + E_T> 22 GeV	> 20 GeV
incl. dimuon	Z, Z', SUSY, ...	> 10 GeV	> 4 GeV

As collider peak luminosity increases

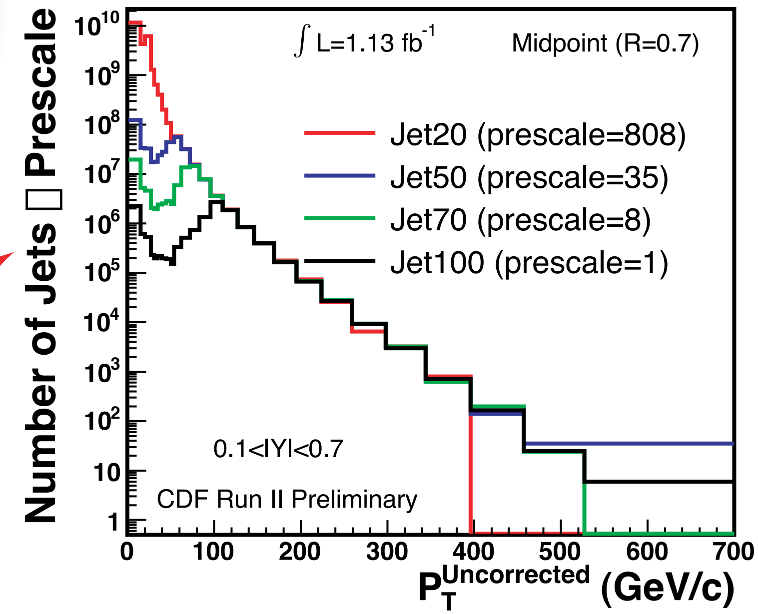
- smarter triggers, tighter cuts, increased prescales, some triggers may be turned off
- pay attention or "your" trigger may become ineffective or disappear!

Wonders of Prescales

- CDF, observed number of jets



- CDF, corrected for prescales

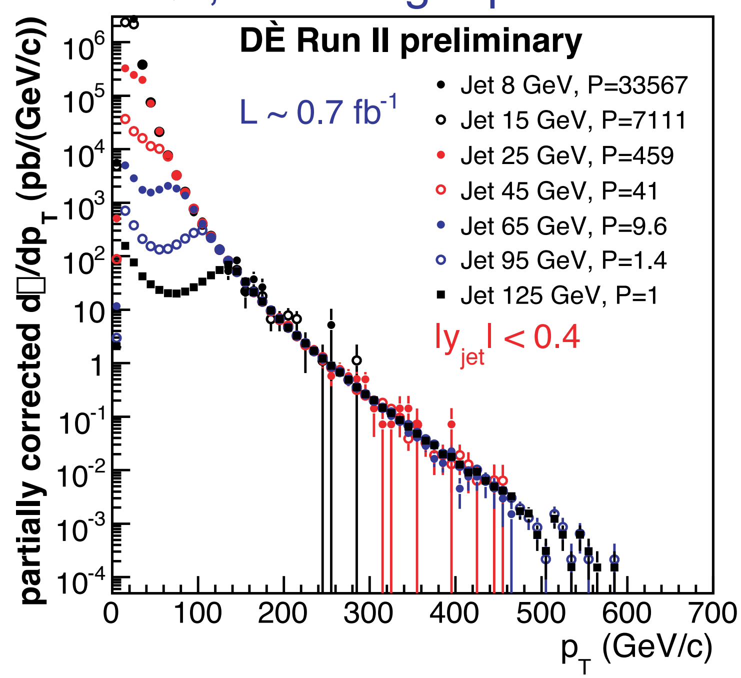


Prescale Factor

$$N_{\text{prod}} = \frac{p_i \cdot N_{\text{obs}}}{\epsilon}$$

"Rebuilding"

- DØ, even larger prescales



Triggering & Analysis

Measuring trigger efficiency; level of knowledge depends on analysis

- Need ϵ_{trig} shape and absolute value,

  e.g., for $\sigma, \dfrac{d\sigma}{dp_T}, \dfrac{d\sigma}{dx}$

- Need ϵ_{trig} shape (or bias caused by it to correct)

  e.g., for measuring mass, lifetime, $\dfrac{1}{N}\dfrac{dN}{dp_T}$

- Just need triggers that don't create a bias, but still need to check level of bias

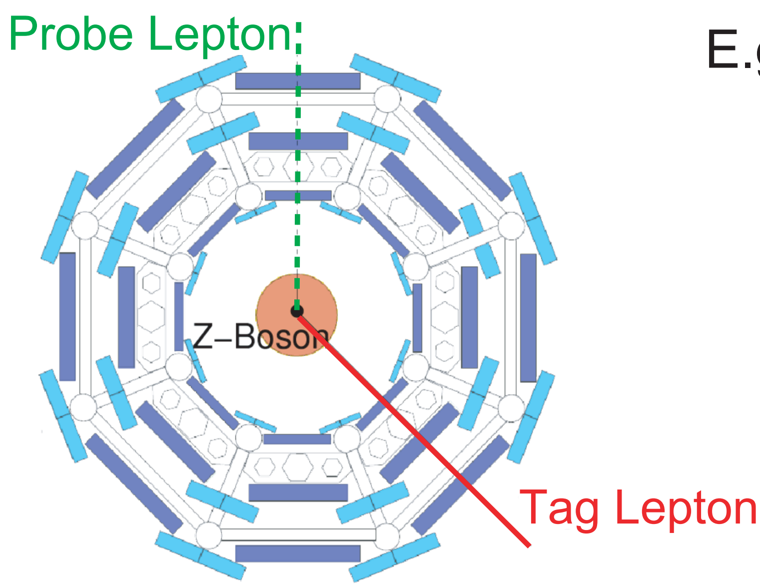
  e.g., for measuring asymmetry, lifetime (e.g., remove triggers involving impact parameter)

Could use a trigger simulator, but data-driven methods always preferred

- Tag & probe methods
- "Bootstrapping"
- Orthogonal triggers
- Reference measurements

Trigger Efficiency

Tag-and-Probe Method



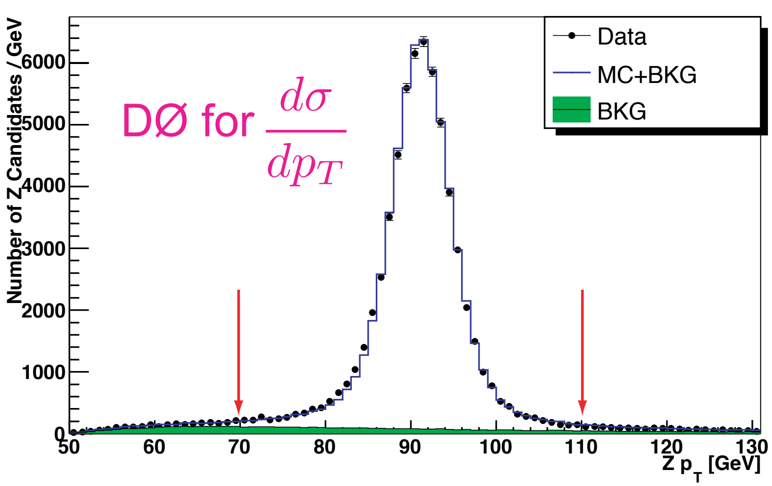
E.g.: lepton trigger efficiency using $Z^0 \to \ell^+\ell^-$

Tag Lepton

- Event triggered by tag electron or muon or tau
- Require some minimum p_T (e.g., > 20 GeV)

Probe Lepton (unbiased w.r.t. tag selection)

- Inv. mass window around $M(\ell^+\ell^-) \approx M_Z$
- Count how often probe lepton fires the lepton trigger

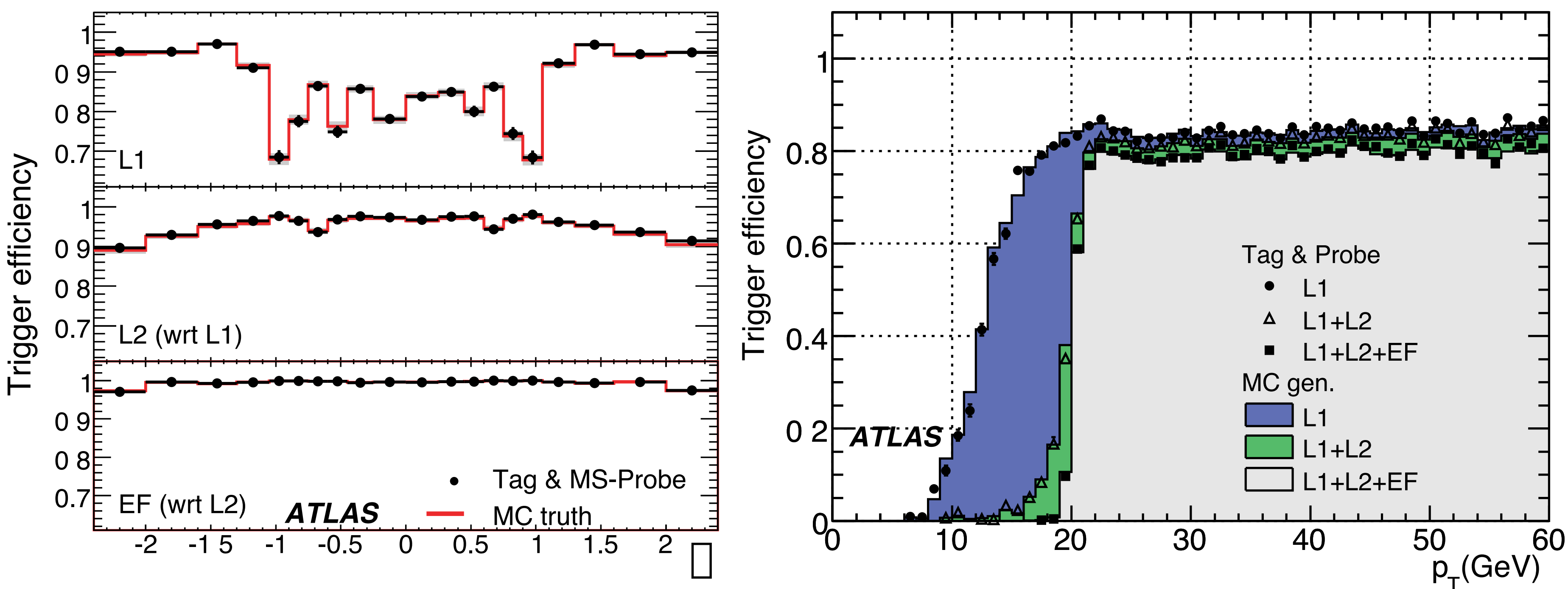


Enough statistics? Can do in bins of your favorite kinematic/geometric variable: p_T, η, ϕ

Clever variations: event counts in single leptons and di-lepton triggers

Trigger Efficiency

Tag-and-Probe Method



- Corresponds to statistics of 50 pb^{-1} [CERN-OPEN-2008-020]

Trigger Efficiency

Orthogonal Trigger

- Use triggers considering information independent of the trigger for which you want the efficiency

  e.g., use calorimeter triggers to create an unbiased sample to test a muon trigger

  or vice-versa...

  e.g., use a muon trigger plus close track activity to create a sample to measure jet calorimeter trigger efficiency

  This sample will be biased towards more heavy-flavor jets (from *b*-hadron semileptonic decay) than light-quark jets; may be what you want!

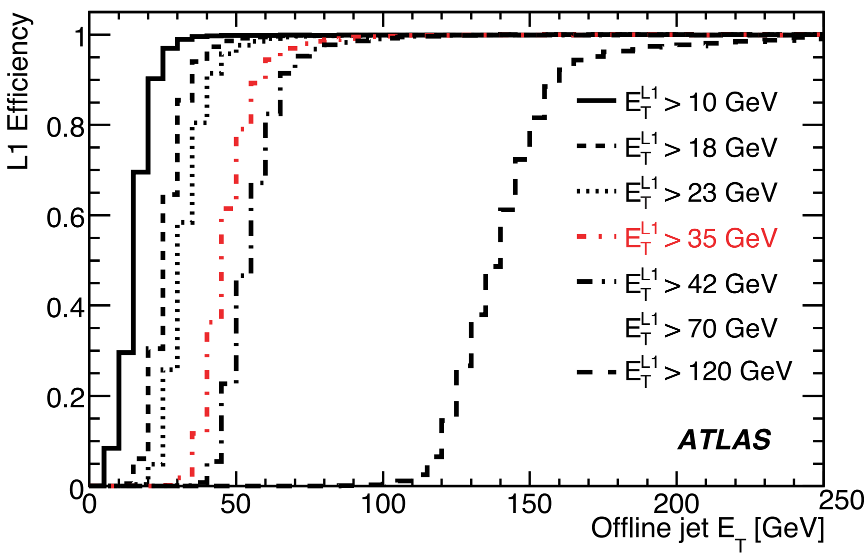
  or it is a possible pitfall if not what you want, and measure the incorrect trigger efficiency

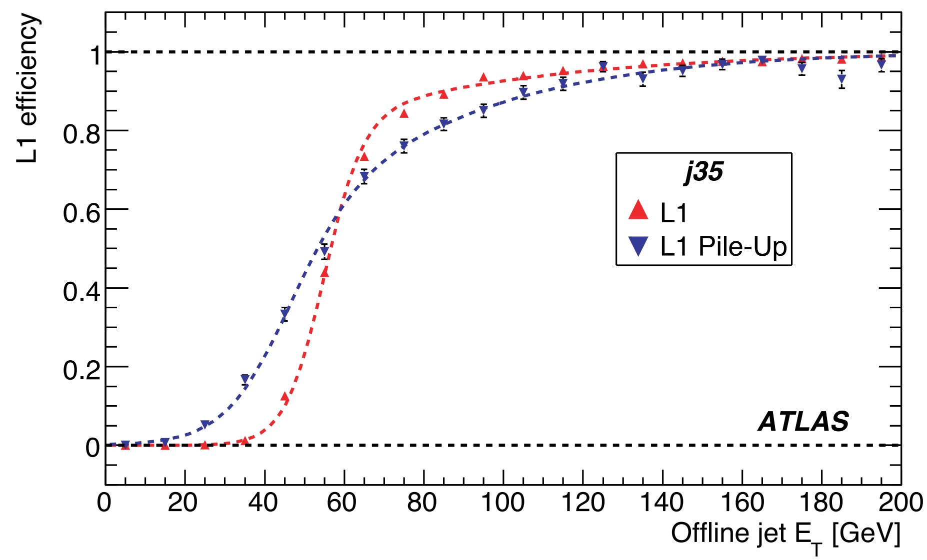
Use a range of different methods and/or samples to measure trigger efficiency;
can use spread to estimate systematic uncertainty

Trigger Efficiency

"*Bootstrapping*"

Jet energy resolution worse ⟶





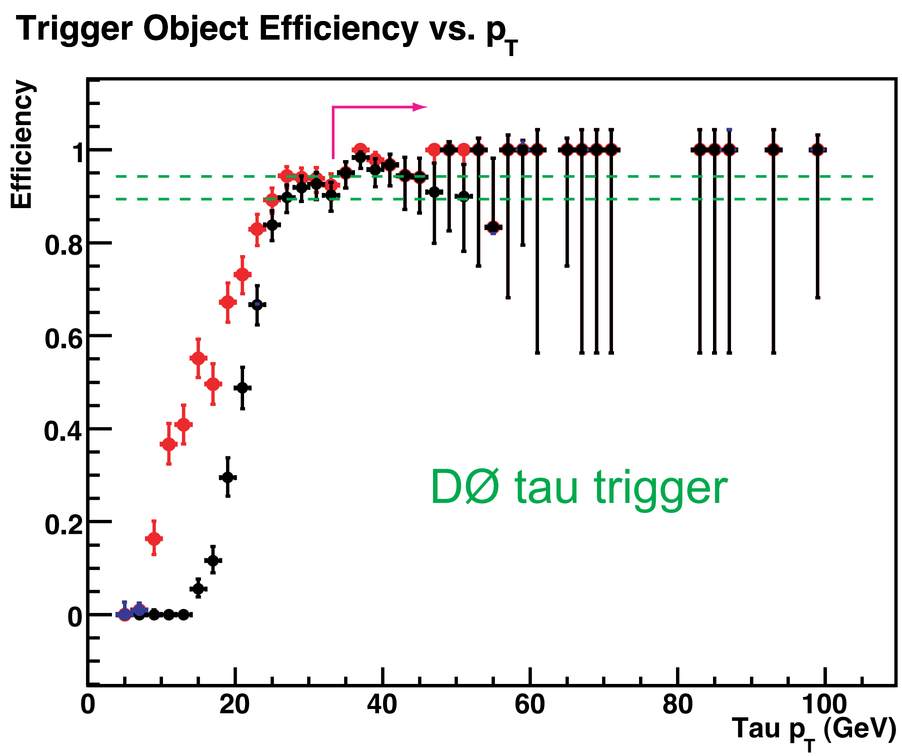
- Use less restrictive trigger sample to determine efficiency of more restrictive one, i.e., E_T > 10 GeV trigger w.r.t. minimum bias sample, E_T > 35 GeV w.r.t. E_T > 10 GeV trigger sample, etc.

- Potential worry: trigger efficiencies can depend on a lot of parameters, e.g., pile-up events, inst. luminosity

Aside: "zero bias trigger" would be a random trigger on bunch crossing, whether there is an interaction in that bunch crossing or not: useful for determining noise in calorimeters, etc.!

Trigger Efficiency

Be "trigger aware"

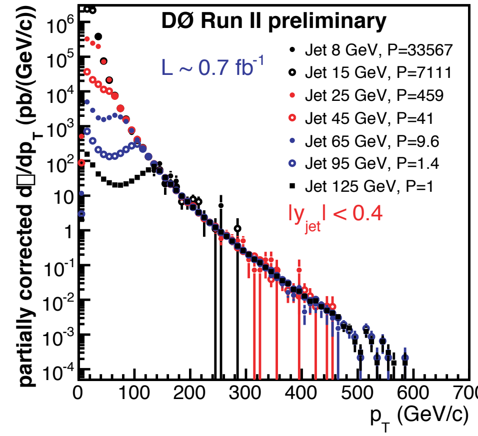


Red points are a *combined* trigger

- Try to avoid the messy "turn-on" region in offline criteria; uncertainties in this region may not be worth it

Combining Triggers

Why?



Increase number of signal events *or* cover more of phase space

- Different energies, phase space (already seen)
- Different subdetectors (e.g., barrel & endcap)
- Different signals (e.g., muons *or* jets)

Generally three different methods:

Excellent reference: arXiv:0901.4118 for detailed weighting formulae

- Division method (simple, may be sufficient):

  One trigger line per distinct (divided) phase space region

- Exclusion method (split data according to trigger lines and prescale factors)

  Start with a *fully efficient trigger combination* (FETC)

  (i.e., each event was taken by at least one raw trigger, and in each part of phase space at least one trigger is fully efficient)

  Choose trigger line with smallest prescale factor for which the "raw trigger" fires

- Inclusion method (can be complicated, but best)

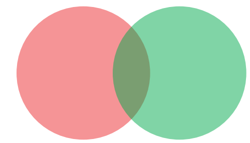
  At least one of a list of trigger lines fires

Combining Triggers

Inclusion Method

At least one of a list of trigger lines fires

E.g., for combining two trigger lines, then probability:



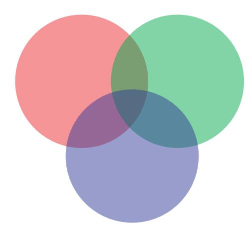
$$Pr_{\text{tot}}(\text{evt}) = Pr_1(\text{evt}) + Pr_2(\text{evt}) - \underbrace{Pr_1(\text{evt}) \cdot Pr_2(\text{evt})}_{\text{Overlap prob.}} \cdot Pr_{2|1}(\text{evt})$$

and if efficiency correlations

Conditional prob.

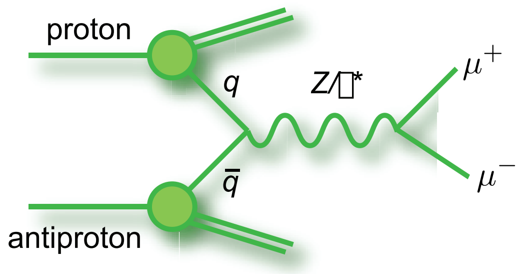
- Instrumental
  - common trigger element with inefficiency
  - common electronics
- Physics
  - e.g., $E_{\text{jet}} > E_0$ may be correlated
    $N_{\text{trk}} > N_0$

Can get increasingly complex as number of trigger lines increases,
can be solved recursively



Cross Section: An Example

A "standard candle", usual to measure at start-up of a new machine...



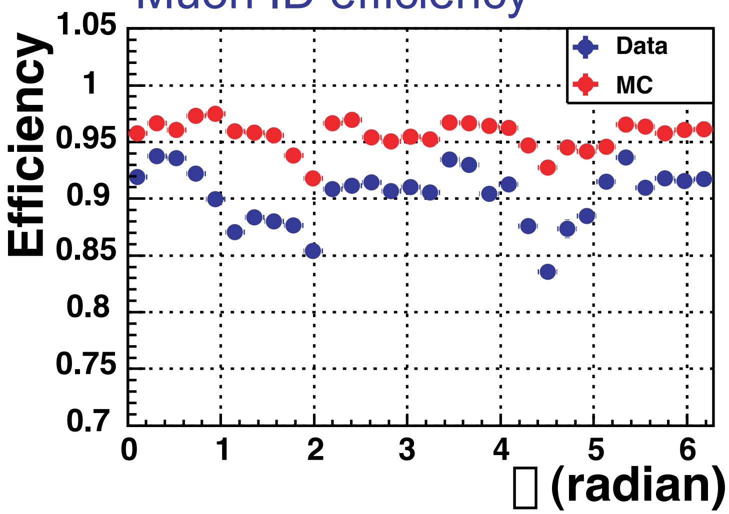
$$\sigma(p\bar{p} \to Z^0 X) \cdot \mathcal{B}(Z^0 \to \mu^+\mu^-)$$

- Single or di-muon triggers
- Two oppositely charged, identified muons, isolated
- $p_T(\mu) > 20\,\mathrm{GeV}$
- $70 < M(\mu^+\mu^-) < 110\,\mathrm{GeV}$

$$\sigma = \frac{N_{\mathrm{obs}} - N_{\mathrm{backg}}}{\epsilon \cdot \int \mathcal{L} dt}$$

$$\epsilon_{(\mathrm{total})} = \epsilon_{\mathrm{trig}} \cdot \epsilon_{\mathrm{rec}} \cdot \epsilon_{\mathrm{ID}} \cdot \epsilon_{\mathrm{kin}}$$

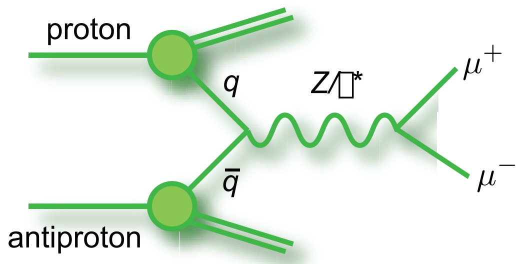
Muon ID efficiency



- Muon identification: see *Particle Identification*, Olav Ullaland
- For efficiency, in data, can use reference muon samples and/or "tag & probe" method using Z^0, $\Upsilon(4S)$, $J/\psi \to \mu^+\mu^-$

Cross Section: An Example

A "standard candle", usual to measure at start-up of a new machine...

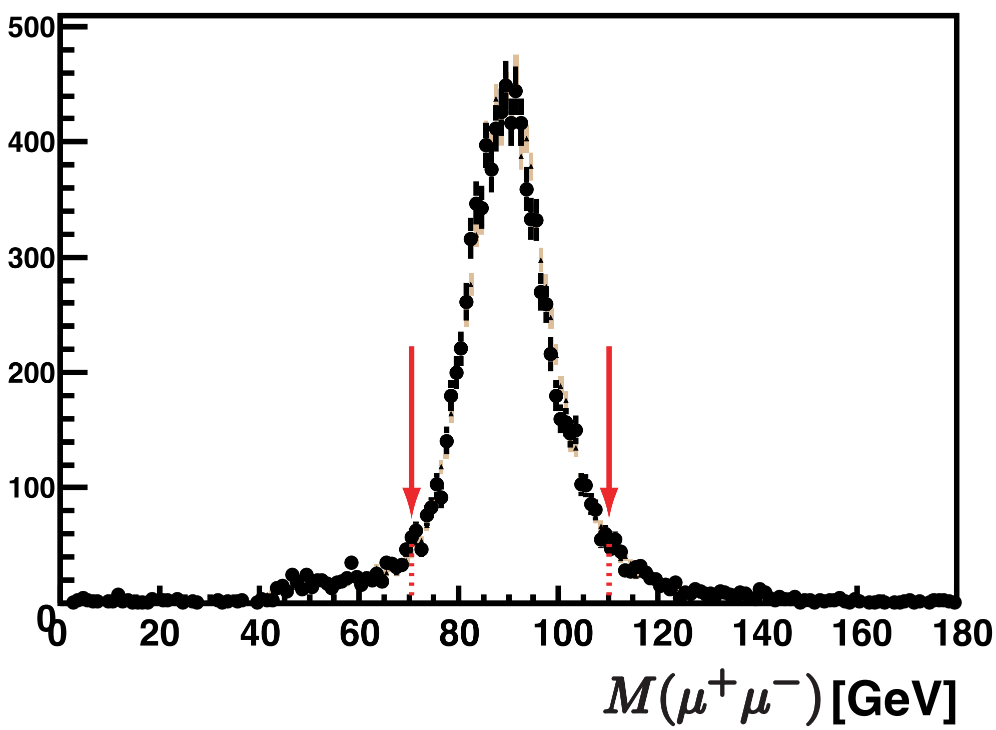


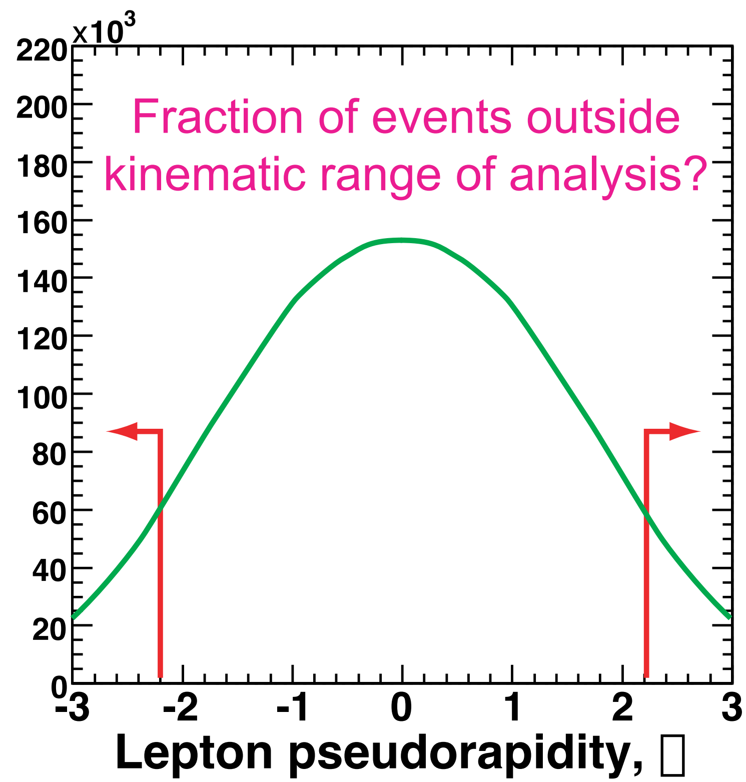
$$\sigma(p\bar{p} \to Z^0 X) \cdot \mathcal{B}(Z^0 \to \mu^+\mu^-)$$

$$\epsilon_{(\mathrm{total})} = \epsilon_{\mathrm{trig}} \cdot \epsilon_{\mathrm{rec}} \cdot \epsilon_{\mathrm{ID}} \cdot \boldsymbol{\epsilon_{\mathrm{kin}}}$$

- For kinematic & geometric efficiency/acceptance

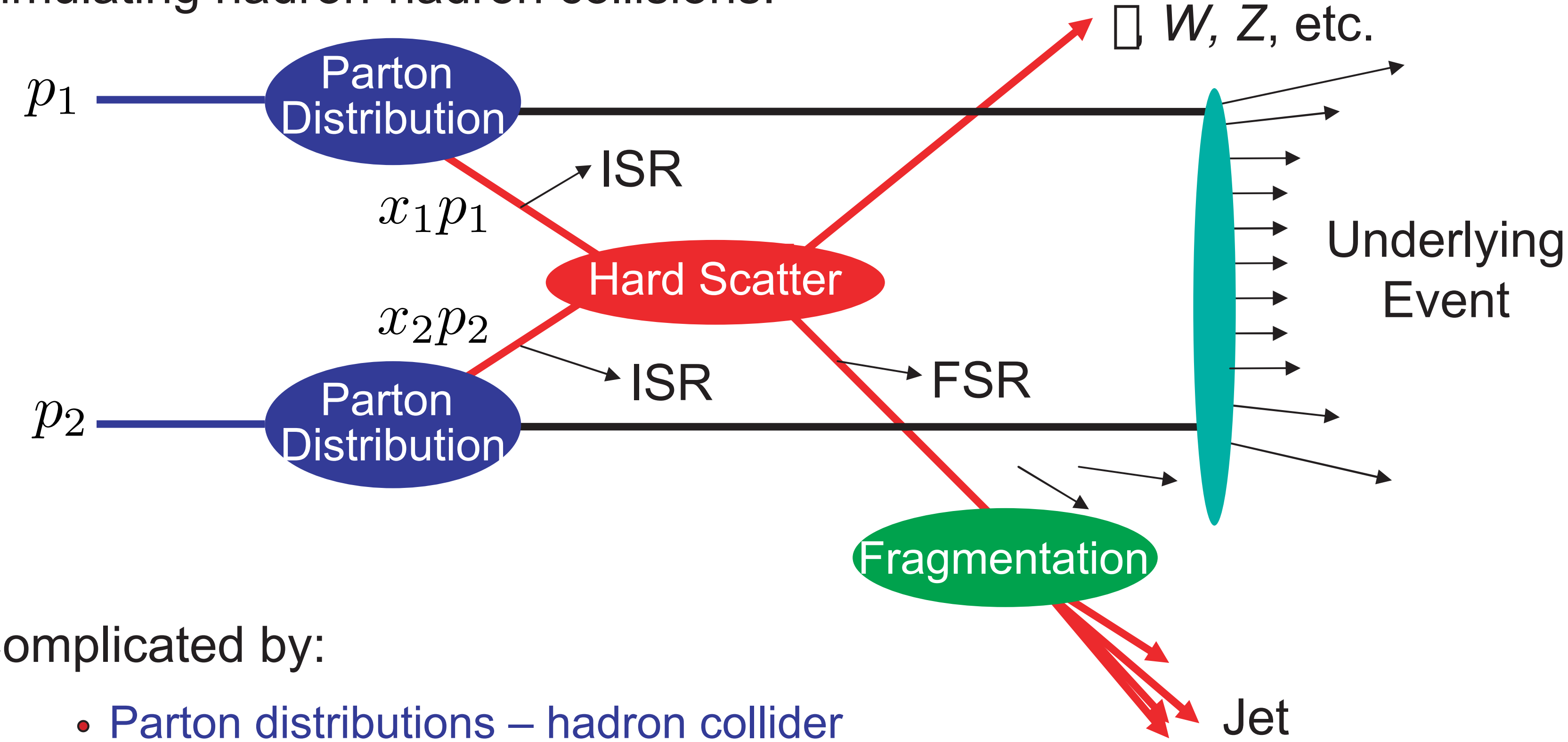
→ Monte Carlo simulation!





Monte Carlo Simulation

Simulating hadron-hadron collisions:



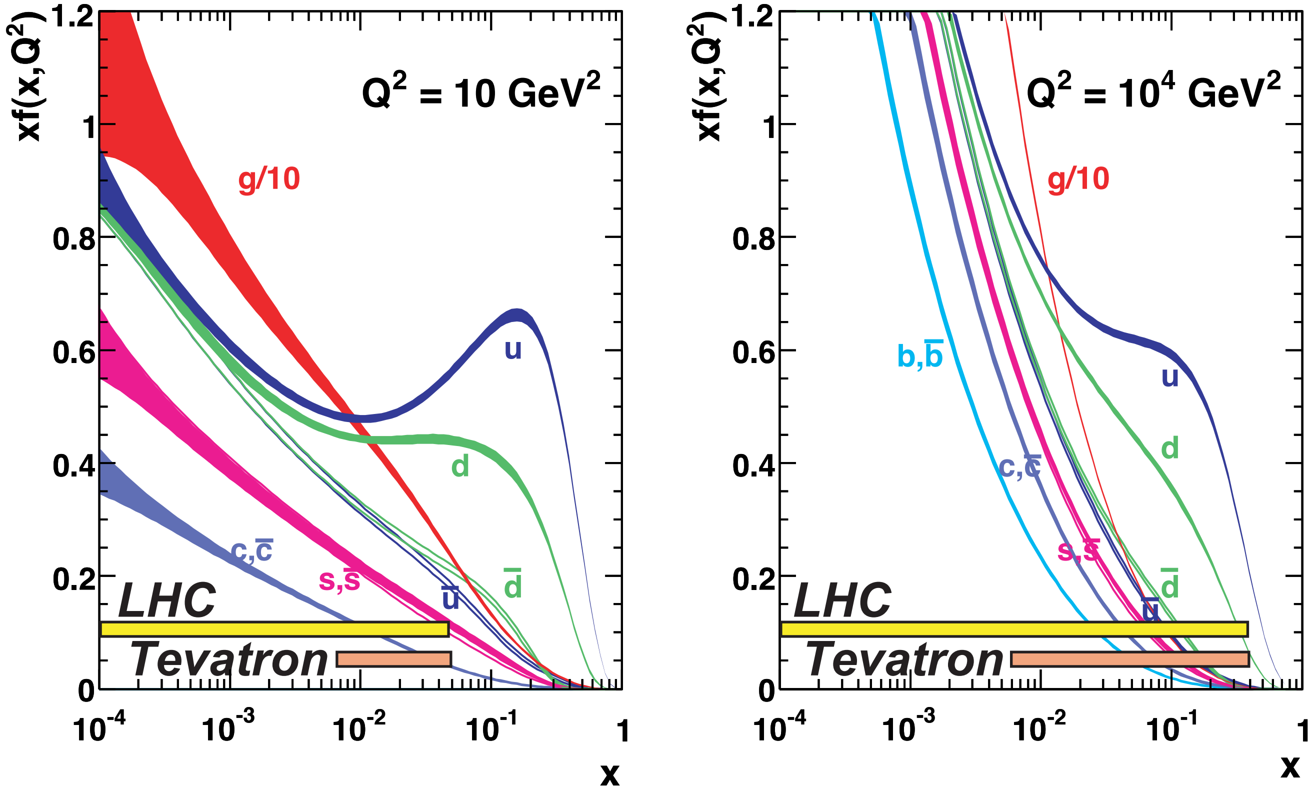
Complicated by:

- Parton distributions – hadron collider is really a "broad-band" quark & gluon collider
- Both initial and final state radiation (ISR & FSR) can have color, i.e., radiate gluons (soft jets)
- Underlying event due to proton (anti-proton) remnants

Monte Carlo Simulation

Both acceptance/efficiency *and* cross sections sensitive to PDF's

One example set and uncertainties: **MSTW 2008 NLO PDFs (68% C.L.)**



$M \approx \sqrt{Q^2}$

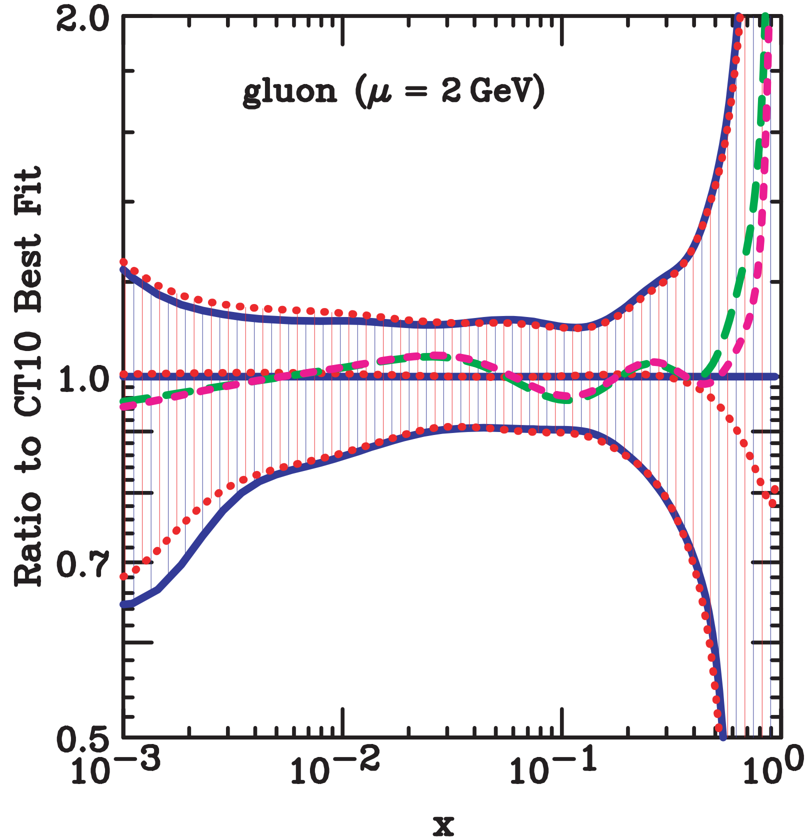
LHC essentially a gluon-gluon collider

- Can lead to some sizeable systematic uncertainties!!
- Other sets: CT10 (CTEQ6.6), NNPDF2.0, HERAPDF, ADKM09, GJR08
- Can access most under common interface: LHAPDF (Les Houches Accord)

Monte Carlo Simulation

Implementing PDF uncertainties?

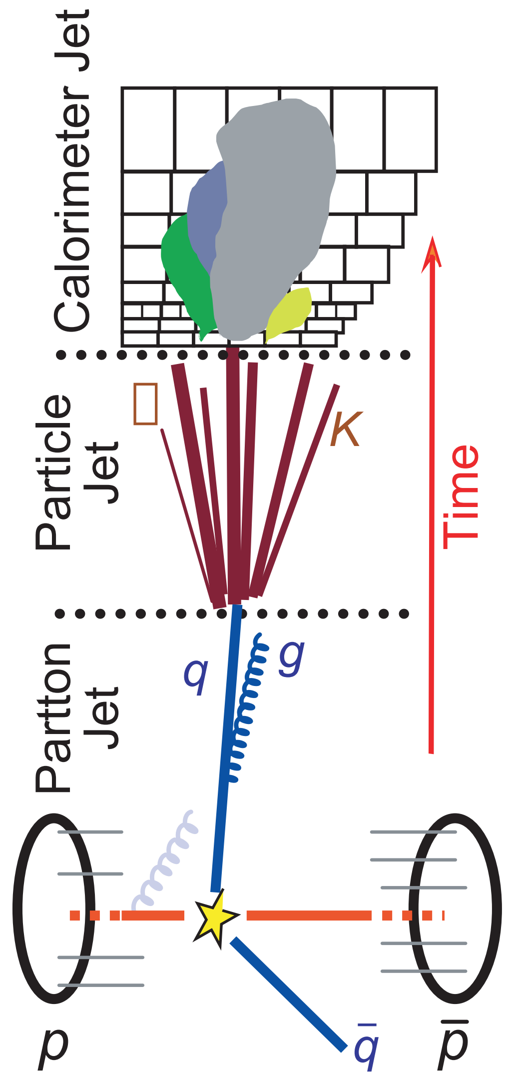
Each set defines "eigenvectors" of global fit parameters to the input data, each of which represents an acceptable fit to the 1-σ level



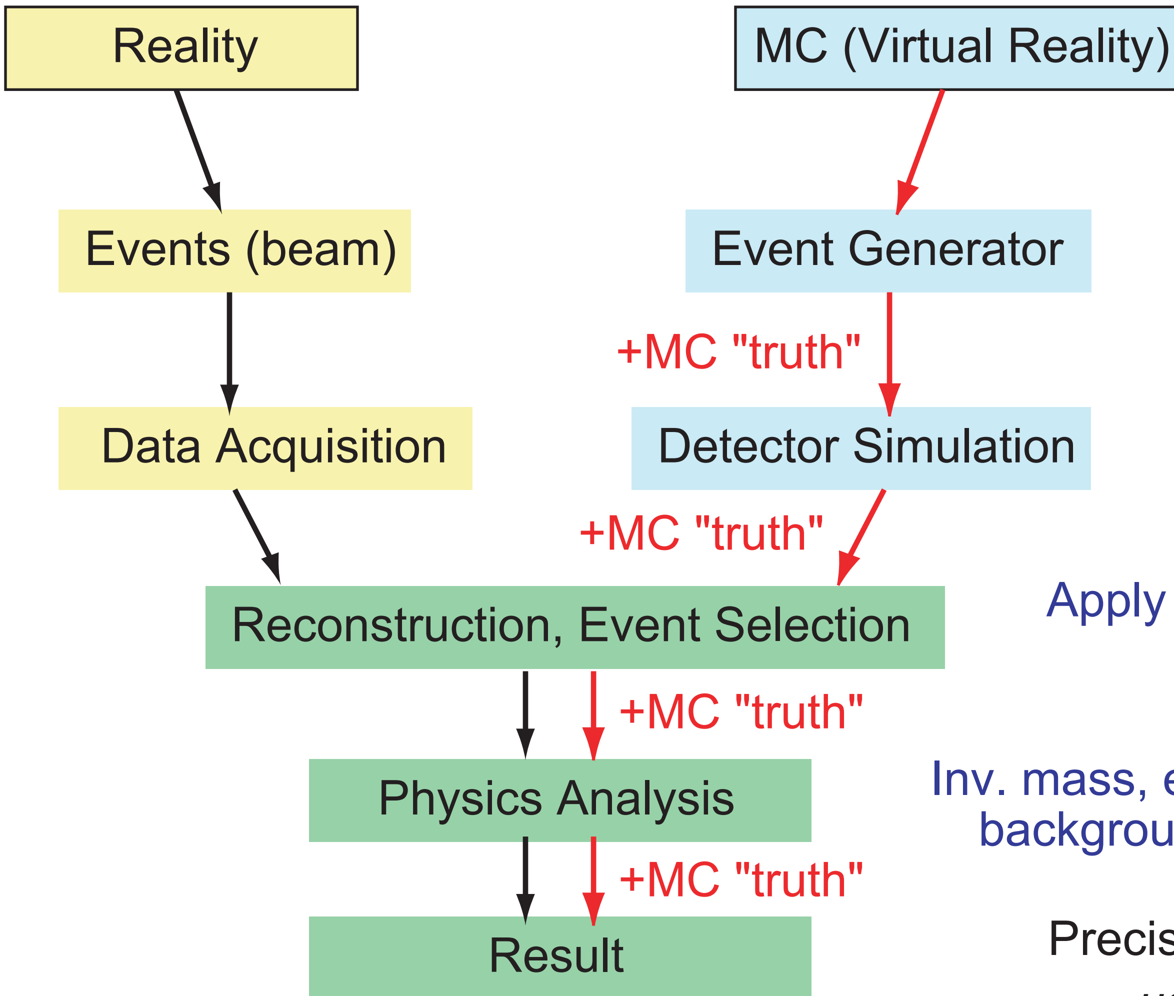
e.g., there are 52 eigenvectors for CT10 that also cover variations in α_s and scale μ

Monte Carlo Simulation

- A "Monte Carlo" is a Fortran or C++ program that generates events
- Events vary from one to the next (random numbers): expect to reproduce the average behavior and fluctuations of real data
- **Event Generators** may be:
  - Parton level
    - Parton Distribution functions
    - Hard interaction matrix element
  - *May* also handle
    - Initial & final state radiation
    - Underlying event
    - Hadronization & decays
- **Detector Simulation** in a separate program
    - GEANT by far most commonly used



Monte Carlo Simulation



Generates 4-vectors for the particles, resonances, ang. dist., decays, etc. (PYTHIA, HERWIG, ALPGEN, Sherpa...)

Generates detector relevant quantities (GEANT 4)

Apply boundary conditions
Acceptance

Inv. mass, efficiency, purity backgrounds, *any* dist.

Precision $\approx 1/\sqrt{N}$

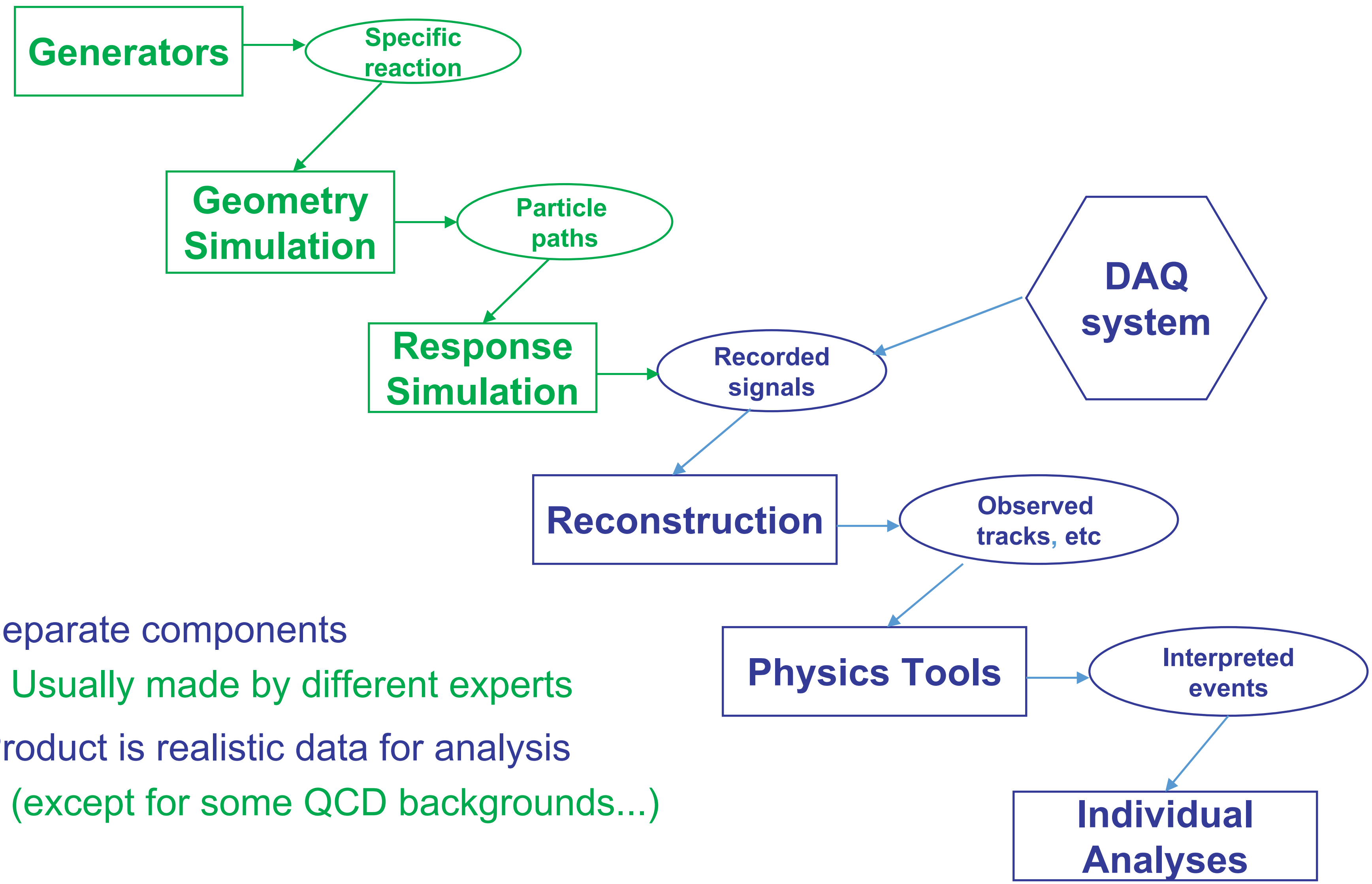
usually:

$$N^{\mathrm{MC}}_{\mathrm{signal}} \gg N^{\mathrm{Data}}_{\mathrm{signal}}$$

$$N^{\mathrm{MC}}_{\mathrm{backg}} \ll N^{\mathrm{Data}}_{\mathrm{backg}}$$ (QCD!)

Monte Carlo Simulation

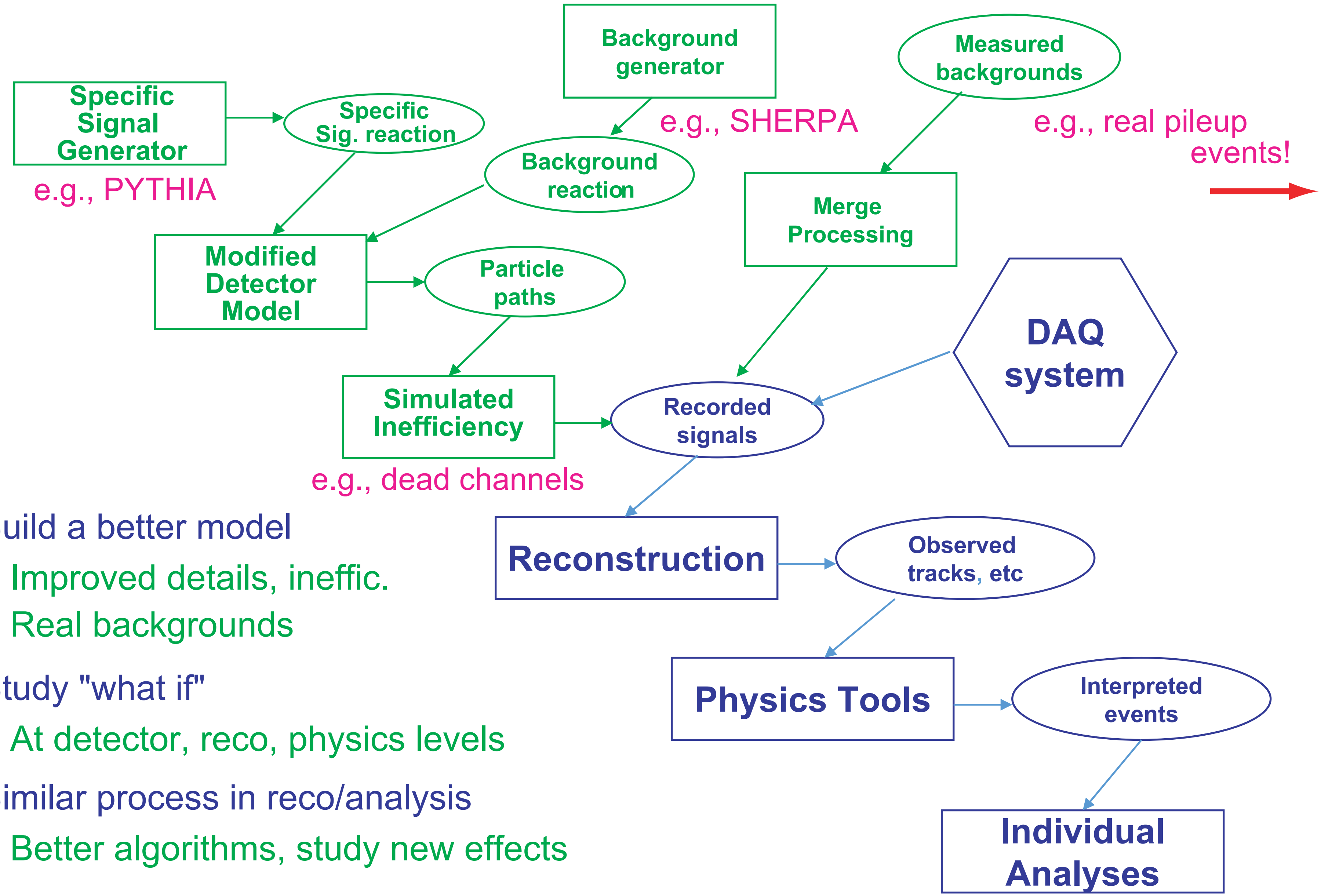
Details



- Separate components
  Usually made by different experts
- Product is realistic data for analysis
  (except for some QCD backgrounds...)

Monte Carlo Simulation

Improve, iterate

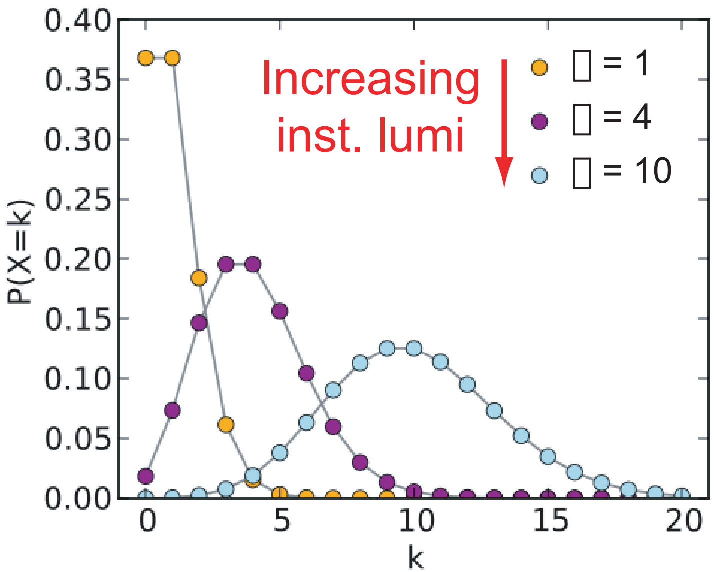


- Build a better model
  - Improved details, ineffic.
  - Real backgrounds
- Study "what if"
  - At detector, reco, physics levels
- Similar process in reco/analysis
  - Better algorithms, study new effects

Monte Carlo Simulation

e.g., overlay/merge real pile-up events on to MC signal or background events

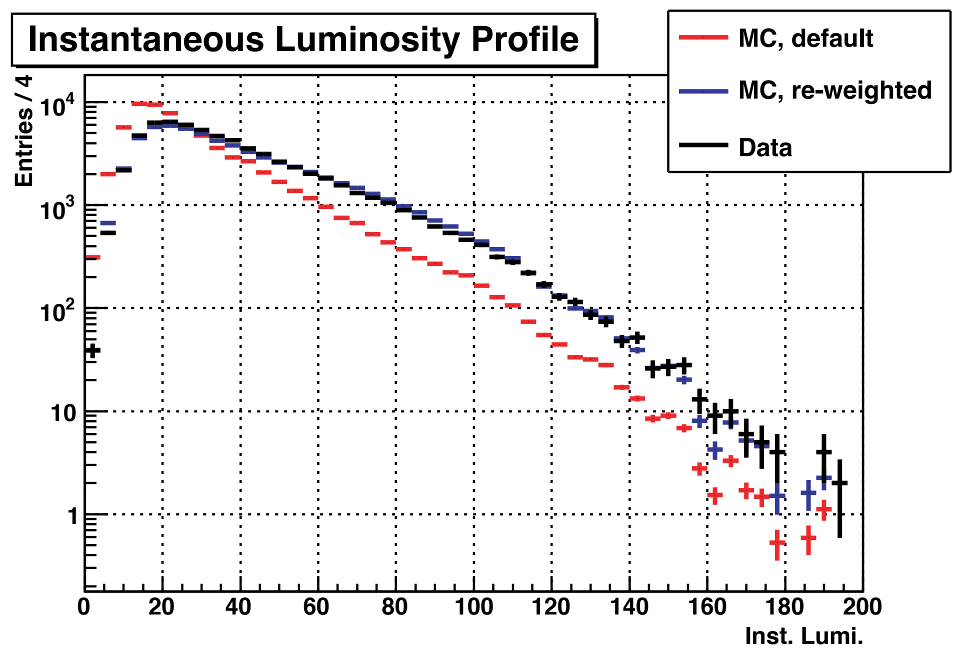
(important for isolation effic., calorimeter activity, tracking performance, triggering, etc.)



- Number of independent pile-up events, *k*, to "overlay" drawn from Poisson dist.,with □ depending on instantaneous luminosity

- If data and MC don't match? Can reweight (within reason)

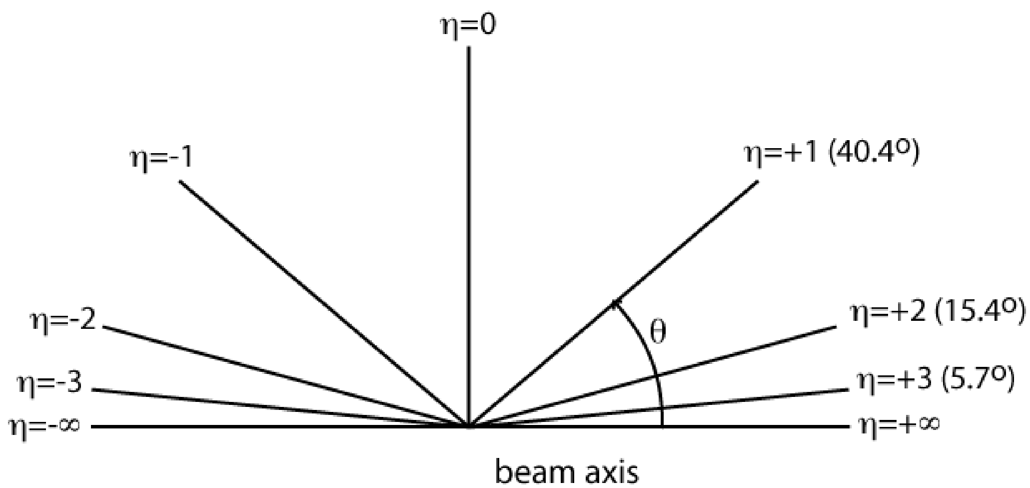
- e.g., to get to match, reweight events with smaller *k* with a weight, *W* < 1, and those with larger *k*, *W* < 1
(e.g., as entered into histogram and entire analysis)

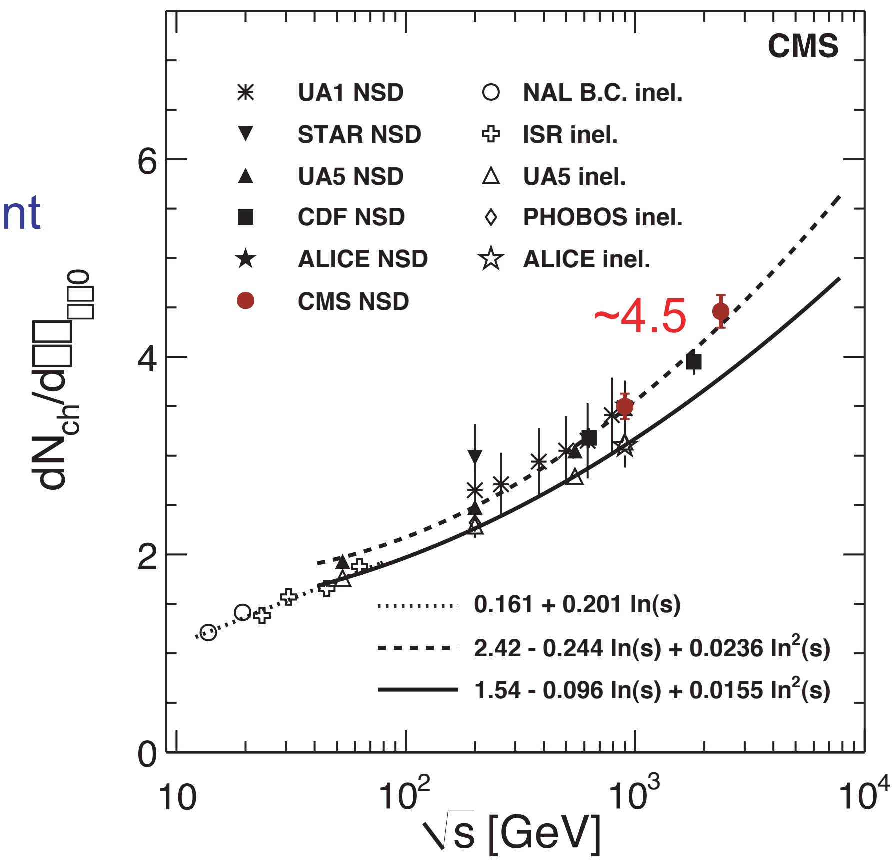


Monte Carlo Simulation

Underlying Event

- Color strings breaking lead to a sort of cloud of soft hadrons in the events
- Often think in terms of the underlying event actually being a min-bias event accompanying the hard collision (or vice versa) – not quite: color reconnection and "beam drag"
- Rule of thumb: number of particles per unit of pseudorapidity is roughly constant...but at what?





~4.5 at <p_T> ~0.5 GeV

Monte Carlo Simulation

Three typical levels of MC simulation:

- Full

  Particle → Energy Deposit → Detector Response → Electronics → Analog Signal → Digitization

  Time consuming, smaller samples

- "Fast" or parameterized

  Intelligently smeared 4-vectors, effiiciencies, noise (from data and full MC)

  And/or calorimeter shower libraries

  Larger samples

- Toy

  Only throw from the handful of prob. dist. functions that you care about (with correlations)

  "Roll your own", usually write (easy in root!) and run yourself

  Crazy-large samples, quickly

  To determine probability of fluctuations, checks for systematic effects, etc..

Monte Carlo Simulation

Points to keep in mind...

Event generators:

- May or may not generate additional jets through parton showering
- May or may not treat spins properly (does it matter to you?)
- May or may not get the cross section correct

  NLO much better than LO, but sometimes no choice

  But can get the "shape" ~right → "*K* factors" = NLO/LO fudge factors

Detector simulation:

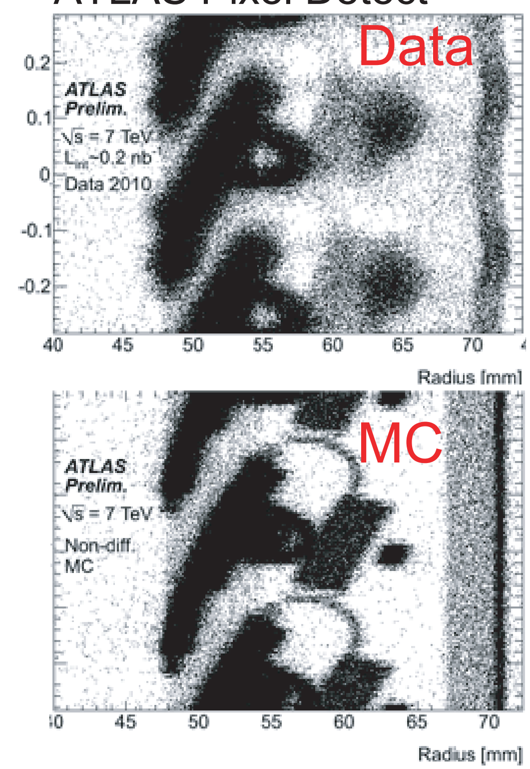
- Your detector simulation is only as good as the geometrical modeling of the detector

  Are all the cables and support infrastructure in place?

  (check with photon conversions & sec. interact.)

- EM showers can be modeled very well (as long as correct material there), but hadronic shower simulation is known to be an imperfect art!

ATLAS Pixel Detect

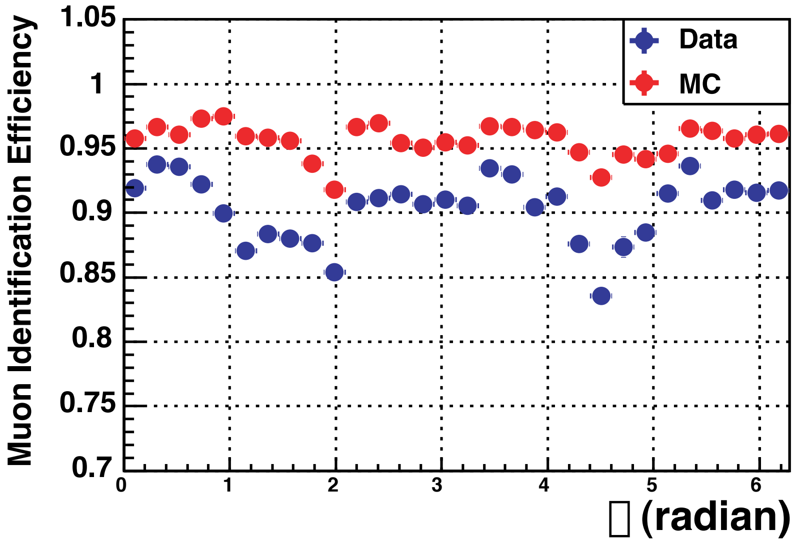


Monte Carlo Simulation

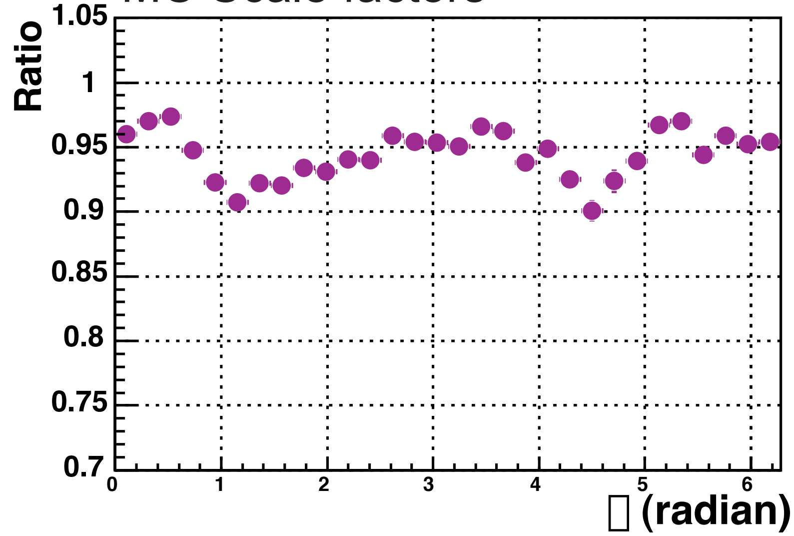
Examples of MC tweaks

Monte Carlo may not be getting every detail

- "Scale factor" corrections to MC determined from data & used consistently



MC Scale factors



For the example, further necessary steps:

- Slight further smearing of MC muon momentum
- Reweighting to NLO (p_T of Z)
- Reweighting of PDF's
- Reweighting luminosity profile

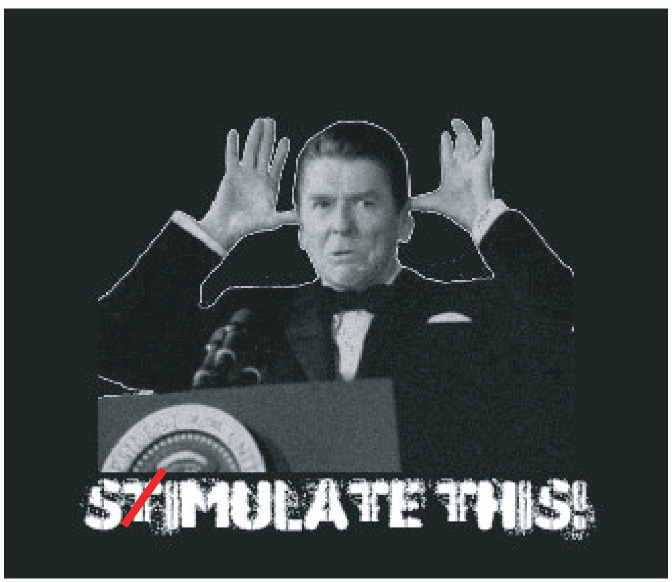
Monte Carlo Simulation

Bottom Line

In general:

- You can usually get the average behavior right
- Don't blindly trust tails of distributions or rare processes!
  - Random numbers may not populate them fully
  - Modeling is not verified at this level
  - The real world is not always Gaussian: ***non-Gaussian tails***

Don't think of simulations so much as absolute predictive tools but as multidimensional parameterizations of knowledge of the detector and SM processes



"Trust but verify" – R. Reagan

"Be suspicious and verify" – RvK

Check your parameterization
in your phase space region of interest!

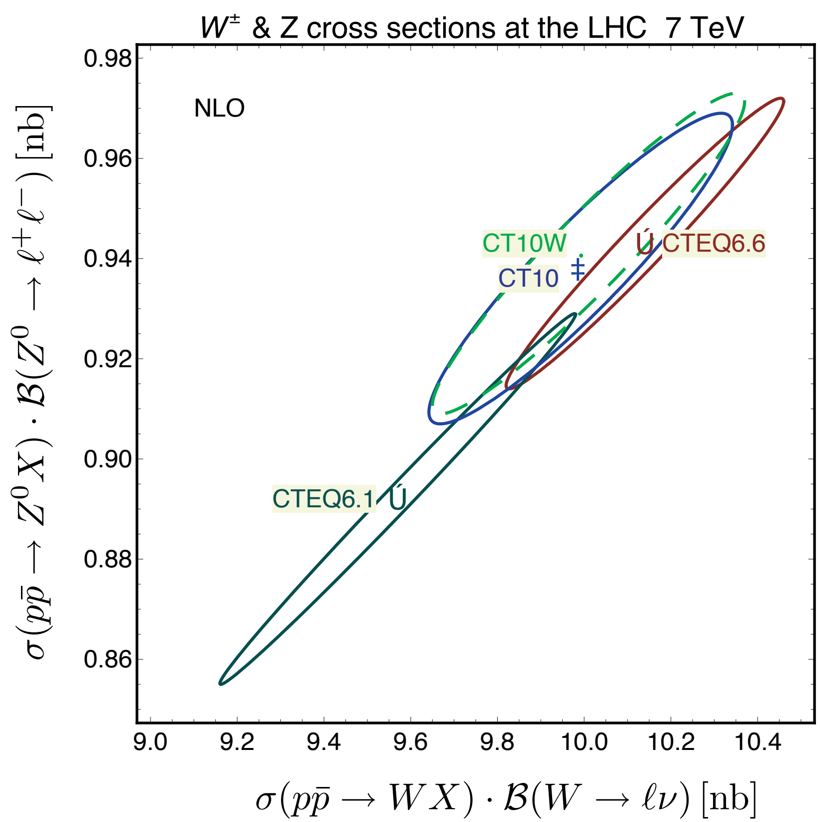
Measuring a Cross Section

Would have all the pieces together, e.g.,

$$\sigma(p\bar{p} \to Z^0 X) \cdot \mathcal{B}(Z^0 \to \mu^+\mu^-) = 265.8 \pm 1.9\,(\text{stat})^{+4.5}_{-5.1}\,(\text{syst}) \pm 16.3\,(\text{lumi})\,\text{pb}$$

Quickly dominated by systematic and luminosity uncertainty;
experimentally, ratios are preferred as luminosity uncertainty could cancel.

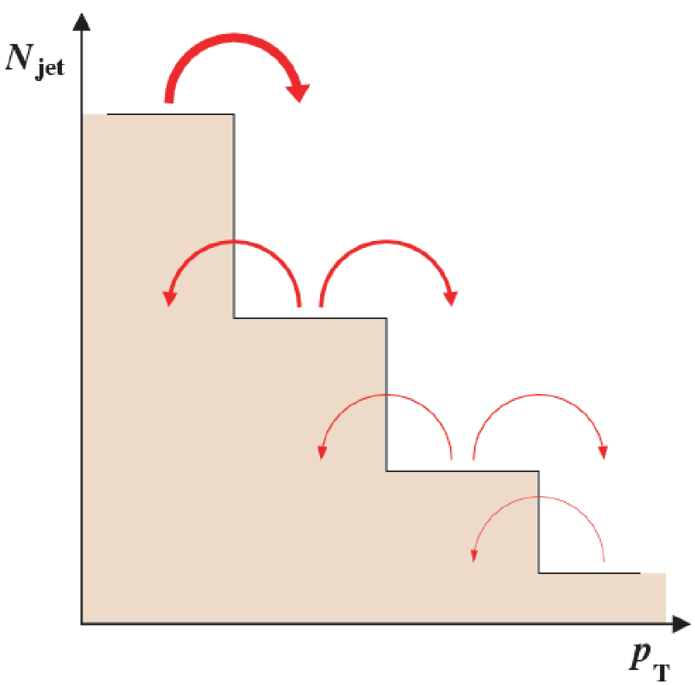
Although:

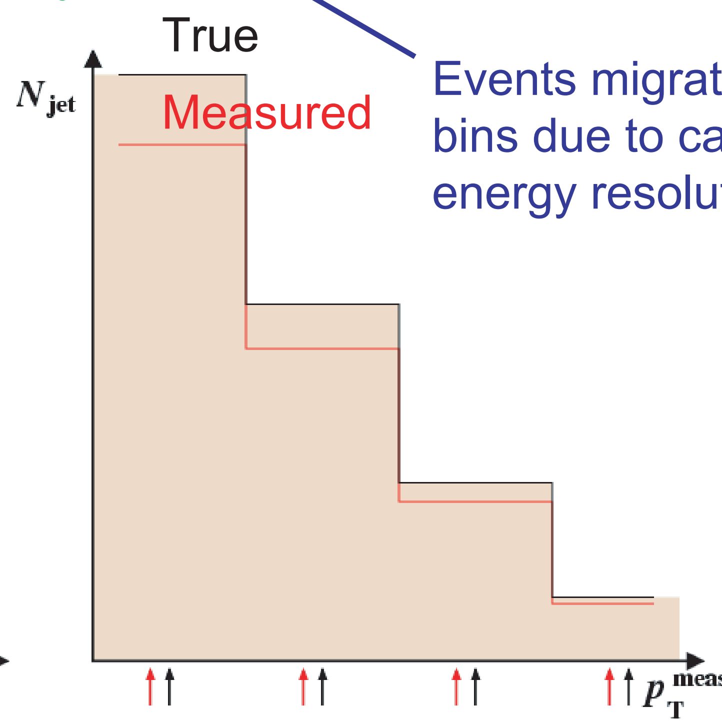


Differential Cross Section

Worry about the shape (particularly steeply falling distribution) and finite resolution:

$$\frac{d^2\sigma}{dp_T dy} = \frac{N}{\epsilon \cdot L \cdot \Delta p_T \Delta y} \cdot C_{\mathrm{smear}} \text{ vs. } p_T$$





Events migrate between bins due to calorimeter energy resolution

We can measure the resolution in data using dijet asymmetry *A*

$$A = \frac{p_{T,1}^{jet} - p_{T,2}^{jet}}{p_{T,1}^{jet} + p_{T,2}^{jet}} \longrightarrow \frac{\sigma_{p_T}}{p_T^{jet}} = \sqrt{2}\sigma_A$$

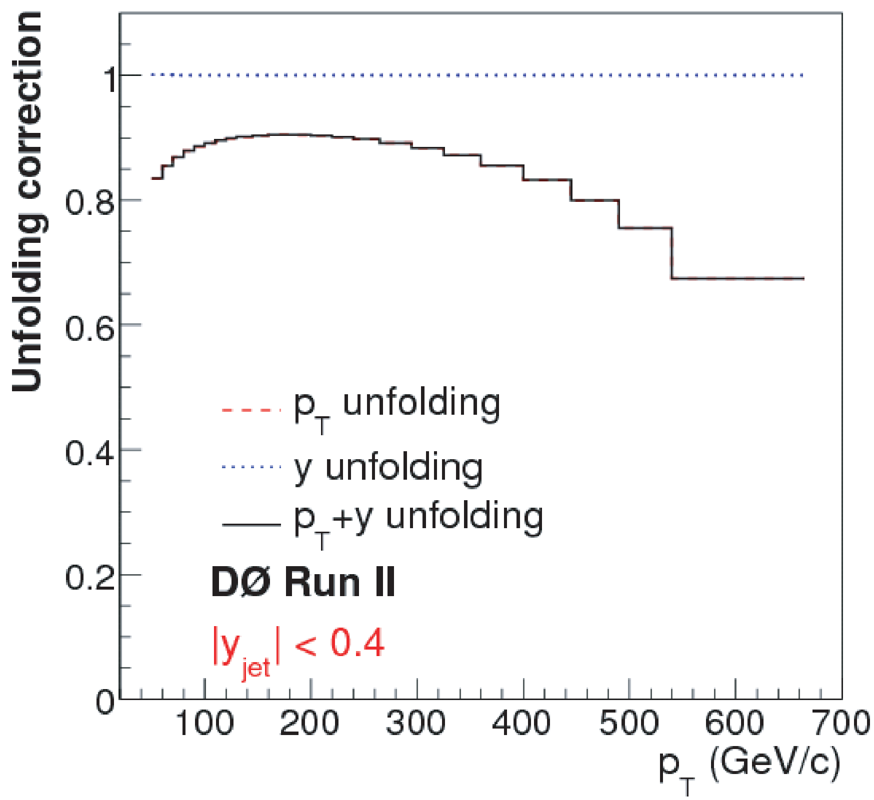
plus lots of corrections

Differential Cross Section

Unfolding

Unfold, using iterative procedure:

- Reasonable MC model (ansatz), smear with resolution
- Fit measurement
- Reweight MC to reflect data measurement; repeat



Works because large statistics, smooth; fluctuations wreck this!

Unfolding

When?

Use unfolding to recover theoretical distribution where

- There is no a-priori parameterisation (otherwise can just fit to function!)
- This is needed for the result and not just comparison with MC
- There is significant bin-to-bin migration of event

Where?

- Traditionally used to extract structure functions
- Dalitz plots: cross-feed between bins due to misreconstruction
- “True” decay momentum distributions

  Theory at parton level, we measure hadrons

  Correct for hadronisation as well as detector effects

How?

- Can sometimes get away with simple iterative procedure
- If low statistics in bins, "spiky", need to smooth $\longrightarrow$ "regularization"
- Packages out there, e.g., RooUnfold, works in root.